

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 16, 2005, 00:16:21 ; Search time 11792 Seconds
(without alignments)
2983.250 Million cell updates/sec

Title: US-10-633-202-2

Perfect score: 4030

Sequence: 1 MAVGSLVHVRVTLILLWFG.....LFCLHVLFKRTSKEDREG 726

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext: 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp
-Q=/cgn2 1/USPTO spool p/US10633202/runat 14062005 141725 13089/app query.fasta_1.903
-DB=GenEmbl -QFMT=fastgap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10633202 @CGN 1 1 5600 @runat 14062005 141725 13089 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	4030	100.0	2181	6	BD086459 SVPH1-26
2	4030	100.0	2810	9	BC025378 Homo sapi
3	4030	100.0	3230	6	AX704773 Sequence
4	4030	100.0	3314	9	AF158643 Homo sapi

C	5	4030	100.0	209483	9	CNS0505TDV	AL357153 Human chr
	6	4004	99.4	2426	6	A99212	A99212 Sequence 13
	7	4004	99.4	2426	6	BD196410	BD196410 Novel met
	8	4004	99.4	2442	6	AX775081	AX775081 Sequence
	9	4004	99.4	2442	9	AF029899	AF029899 Homo sapi
	10	2435	60.4	2169	6	BD137083	BD137083 Testis-sp
	11	2435	60.4	2169	6	AR256989	AR256989 Sequence
	12	2435	60.4	3557	9	AF158644	AF158644 Homo sapi
	13	2425.5	60.2	18756	9	CNS060C84	AL391223 Human chr
	14	2341.5	58.1	2016	6	CQ736276	CQ736276 Sequence
	15	2330	57.8	2622	2	AF158640	AF158640 Homo sapi
	16	2319	57.5	146055	2	AC009607	AC009607 Homo sapi
	17	2319	57.5	206256	9	CNS018GT	AL160191 Human chr
	18	2108.5	52.3	91995	9	AC105914	AC105914 Homo sapi
	19	2102.5	52.2	2301	6	BD267491	BD267491 New metal
	20	2102.5	52.2	2301	9	AF171930	AF171930 Homo sapi
	21	2102.5	52.2	2364	6	BD267492	BD267492 New metal
	22	2102.5	52.2	2364	9	AF171931	AF171931 Homo sapi
	23	2102.5	52.2	2463	6	BD267493	BD267493 New metal
	24	2102.5	52.2	2463	9	AF171929	AF171929 Homo sapi
	25	2102.5	52.2	2958	9	AF134708	AF134708 Homo sapi
C	26	2053	50.9	234832	2	AC127883	AC127883 Rattus no
	27	2050	50.9	2190	10	AF251559	AF251559 Mus muscu
C	28	2050	50.9	174472	2	AC074356	AC074356 Mus muscu
	29	1994.5	49.5	2640	10	AF167403	AF167403 Mus muscu
C	30	1982.5	49.2	190094	2	AC131819	AC131819 Rattus no
C	31	1982.5	49.2	224808	2	AC094882	AC094882 Rattus no
	32	1965	48.8	168871	2	AC016117	AC016117 Homo sapi
	33	1965	48.8	178517	2	AP006246	AP006246 Homo sapi
	34	1952	48.4	4593	9	AF158639	AF158639 Homo sapi
	35	1951	48.4	154065	2	AC084079	AC084079 Homo sapi
	36	1951	48.4	154539	2	AC079187	AC079187 Homo sapi
	37	1951	48.4	186218	9	AP006248	AP006248 Homo sapi
	38	1948.5	48.3	198057	2	AC126072	AC126072 Rattus no
	39	1948.5	48.3	327299	9	AC096502	AC096502 Rattus no
	40	1945	48.3	100000	9	AB020862	AB020862 Homo sapi
	41	1940	48.1	193426	2	AC117729	AC117729 Mus muscu
	42	1934	48.0	2726	10	AY190759	AY190759 Mus muscu
C	43	1902.5	47.2	190187	2	AC121140	AC121140 Mus muscu
	44	1892.5	47.0	2806	10	AF167402	AF167402 Mus muscu
C	45	1846	45.8	256368	2	AC122753	AC122753 Mus muscu

ALIGNMENTS

RESULT 1	BD086459	2181 bp	DNA	linear	PAT 27-AUG-2002
LOCUS	SVPH1-26	DNA and polypeptide.			
DEFINITION	BD086459.1	GI:22632069			
ACCESSION	BD086459				
VERSION	JP 2001521742-A/1.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 2181)				
AUTHORS	Cerretti, D.P.				
TITLE	SVPH1-26 DNA and polypeptide				
JOURNAL	Patent: JP 2001521742-A 1 13-NOV-2001;				
COMMENT	IMMUNEX CORP				
	OS Homo sapiens (human)				
	PN JP 2001521742-A/1				
	PD 13-NOV-2001				
	PF 30-OCT-1998 JP 2000519084				
	PR 30-OCT-1997 US 60/063571				
	PI DOUGLAS PAT CERRETTI				
	PC C12N15/09,C07K16/40,C12N1/19,C12N1/21,C12N5/10,C12N9/64,C12O1/37,C12Q1/68,				
	PC G01N33/68,C12N15/00,C12N5/00				
	CC SVPH1-26 DNA and polypeptide				
	FT Key location/Qualifiers				
	FT source 1. .2181				

```

FEATURES             FT            /organism='Homo sapiens (human)'.
source               1. .2181
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.:          0           Length:      2181
Score:             4030.00      Matches:      726
Percent Similarity: 100.00%     Conservative: 0
Best Local Similarity: 100.00%  Mismatches:   0
Query Match:       100.00%     Indels:      0
DB:                6           Gaps:        0

US-10-633-202-2 (1-726) x BD086459 (1-2181)
Qy  1  MetAlaValGluProLeuValHisIleArgValThrLeuLeuLeuLeuTrpPheGly  20
Db  1  ATGGCAGTGGGTGAGCCCTGGTGACATCAGGGTCACTCTCTGCTGCTCTGGTTGGG  60
Qy  21  MetPheLeuSerIleSerGlyHisSerGlnAlaArgProSerGlnTyrPheThrSerPro  40
Db  61  ATGTTTTTGTCTATTTCTGGCCACTCTCAGCCAGGCCCTCCCACTATTTTCACTTCTCCA  120
Qy  41  GluValValIleProLeuLysValIleSerArgGlyValArgGlyAlaLysAlaProGlyTrp  60
Db  121  GAAGTGGTGATTCCTTTGAAGGTGATCAGCAGGGGACAGAGTGCAAAGGCTCTCTGGATGG  180
Qy  61  LeuSerTyrSerLeuArgPheGlyGlyGlnArgTyrIleValHisMetArgValAsnLys  80
Db  181  CTCTCCTATAGCCTGGCGTTTGGGGGACAGAGATACATTGTCCACATGAGGGTAATAAG  240
Qy  81  LeuLeuPheAlaAlaHisLeuProValPheThrTyrThrGluGlnHisAlaLeuLeuGln  100
Db  241  CTGTTGTTGTGTGCACACCTTCTGTGTTTCACTTACACAGAGCAGCATGCGCTCTCCAG  300
Qy  101  AspGlnProPheIleGlnAspAspCysTyrTyrHisGlyTyrValGluGlyValProGlu  120
Db  301  GATCAGCCCTTCATCCAGGATGACTGCTACTACCATGTTATGTGGAGGGGGTCCCTCGAG  360
Qy  121  SerLeuValAlaLeuSerThrCysSerGlyGlyPheLeuGlyMetLeuGlnIleAsnAsp  140
Db  361  TCCTTTGTTGTCCTTAGTACCTGTTCTGGGGCTTCTTGGAAATGCTACAGATAAATGAC  420
Qy  141  LeuValTyrGluIleLysProIleSerValSerAlaThrPheGluHisLeuValTyrLys  160
Db  421  CTTGTTTTATGAAATCAAGCCAAATAGTGTCTTGCCACATTTGAAACACCTAGTATATAAG  480
Qy  161  IleAspSerAspAspThrGlnPheProProMetArgCysGlyLeuThrGluLysIle  180
Db  481  ATAGACAGTGTATGATACACAGTTTCCACCTATGAGATGTGGTTPAACAGAAGAAATA  540
Qy  181  AlahisGlnMetGluLeuGlnLeuSerTyrAsnPheThrLeuLysGlnSerSerPheVal  200
Db  541  GCACACAGATGGAGTTGCAATTTGTCATATAAATTTCTACTCTGAAGCAAAAGTTCTTTTGTG  600
Qy  201  GlyTrpTrpThrHisGlnArgPheValGluLeuValValValValAspAsnIleArgTyr  220
Db  601  GGCTGGTGGACCCATCAGCGTTTGTGGAGTGTGATGTAATTTTGAATGTAATATATAGATAT  660
Qy  221  LeuPheSerGlnSerAsnAlaThrThrValGlnHisGluValPheAsnValValAsnIle  240
Db  661  CTTTTTCTCTCAAGTAAATGCAACACAGTGCAGCATGAAGTATTTAAAGTTTGCATATATA  720
Qy  241  ValAspSerPheTyrHisProLeuGluValAspValIleLeuThrGlyIleAspIleTrp  260
Db  721  GTGGATTCCTTCTATCATCTCTTGGAGGTTGATGTAATTTTGAATGTAATATATAG  780
Qy  261  ThrIleSerAsnProLeuProThrSerGlyAspLeuAspAsnValLeuGluAspPheSer  280
Db  781  ACTGCATCAATCCACTTCTTACCAGTGGAGACCTAGATAATGTTTTTAGAGGACTTTTCT  840

281  IleTrpLysAsnTyrAsnLeuLeuAsnArgLeuGlnHisAspValAlahisLeuPheIle  300
Db  841  ATTTGGAGAAATATATACCTTATATATCGACTACACATGATGTGTGCACATCTTTTCATA  900
Qy  301  LysAspThrGlnGlyMetLysLeuGlyValAlaTyrValLysGlyIleCysGlnAsnPro  320
Db  901  AAAGACACACAGGCATGAAGCTTGGTGTCCCTATGTTAAAGAAATATGCCAGAATCCT  960
Qy  321  PheAsnThrGlyValAspValPheGluAspAsnArgLeuValValPheAlaIleThrLeu  340
Db  961  TTTAATACTGCGAGTTGATGTTTGAAGACAAACAGGTGGTGGTCTGTTTGAATTTACTTTG  1020
Qy  341  GlyHisGluLeuGlyHisAsnLeuGlyMetGlnHisAspThrGlnTyrCysValCysGlu  360
Db  1021  GGCCACGAGCTTGGTCTATATTTGGGTATGCAACATGACACCCAGTGGTGTGTGGGAG  1080
Qy  361  LeuGlnTrpCysIleMetHisAlaTyrArgLysValThrThrLysPheSerAsnCysSer  380
Db  1081  CTACAGTGGTGCATATATGCTATAGAAAGGTGACAACTAAATTTAGCAACTGCAGT  1140
Qy  381  TyrAlaGlnTyrTrpAspSerThrIleSerSerGlyLeuCysIleGlnProProTyr  400
Db  1141  TATGCCCAATATTTGGACACAGTACTATCAGTAGTGGATATGTATTCAACCGCTCCATAT  1200
Qy  401  ProGlyAsnIlePheArgLeuLysTyrCysGlyAsnLeuValValGluGluGlyGluGlu  420
Db  1201  CCAGGGATATATTTAGACTGAGTACTGTGGAACTAGTGGTTCAGAGAGGGAGGAA  1260
Qy  421  CysAspCysGlyThrIleArgGlnCysAlaLysAspProCysCysLeuLeuAsnCysThr  440
Db  1261  TGTGACTGTGGAACCATACGGCAGTGTGCAAAAGATCCCTGTTGTTCTGTTAAACTGTACT  1320
Qy  441  LeuHisProGlyAlaAlaCysAlaPheGlyIleCysCysLysAspCysLysPheLeuPro  460
Db  1321  CTACATCTCTGGGGCTCTGCTGCTTTTGGAAATATTTGCAAGACTGCAAAATTTCTGCCA  1380
Qy  461  SerGlyThrLeuCysArgGlnGlnValGlyGluCysAspLeuProGluTyrCysAsnGly  480
Db  1381  TCAGGAACCTTTATGTAGACAAACNAGTTGGTGAATGTGACCTTCACAGTGGTGCATGGG  1440
Qy  481  ThrSerHisGlnCysProAspAspValTyrValGlnAspGlyIleSerCysAsnValAsn  500
Db  1441  ACATCCCATCAATGCCCAGATGATGTGTATGTCAGGACGGGATCTCTGTTAATGTGAAT  1500
Qy  501  AlaPheCysTyrGluLysThrCysAsnAsnHisAspIleGlnCysLysGluIlePheGly  520
Db  1501  GCCTTCTGCTATGAAAGACGCTGTAATAACCATGATATACAAATGTAAGAGAGATTTTGGC  1560
Qy  521  GlnAspAlaArgSerAlaSerGlnSerCysTyrGlnGluIleAsnThrGlnGlyAsnArg  540
Db  1561  CAGATGCAAGGAGTGCATCTCAGAGTTGCTACCAAGAAATCAACACCCAGGAACCGT  1620
Qy  541  PheGlyHisCysGlyIleValGlyThrThrTyrValLysCysTrpThrProAspIleMet  560
Db  1621  TTCGGTCACTGTGGTATTTAGGCACAAACATATGTAATAATGTTGGACCCCTGATATCATG  1680
Qy  561  CysGlyArgValGlnCysGluAsnValGlyValIleProAsnLeuIleGluHisSerThr  580
Db  1681  TGTGGAGGGTTCCAGTGTGAAATGTTGGAGTAAATTTCCCAATCTGATAGACATTTCTACA  1740
Qy  581  ValGlnGlnPheHisLeuAsnAspThrThrCysTyrTrpGlyThrAspTyrHisLeuGlyMet  600
Db  1741  GTGCAGCAGTTTCACTCTCAATGACACCATCTGCTGGGGCACTGATTTATCATTTAGGATG  1800
Qy  601  AlaIleProAspIleGlyGluValLysAspGlyThrValCysGlyProGluLysIleCys  620
Db  1801  GCTATACCTGATATTTGGTGGAGTGAAGATGGCACAGTATGTGGTCCAGAAAAGATCTGC  1860
Qy  621  IleArgLysCysAlaSerMetValHisLeuSerGlnAlaCysGlnProLysThrCys  640
Db  1861  ATCCGTAAGAAGTGTCCAGTATGGTTTCATCTGTCAAGCCCTGTGAGCCTTAAGACCTGC  1920
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 16, 2005, 00:22:31 ; Search time 8209 Seconds
(without alignments)
3366.386 Million cell updates/sec

Title: US-10-633-202-2
Perfect score: 4030
Sequence: 1 MAVGRLVHVRVTLVLLWFG.....LFLHLVLFKRTSKEDDEG 726

Scoring table:
BLOSUM62
-Q-/cgn2 1/USPTO spool p/US10633202/runat 14062005 141725 13099/app_query.fasta_1.903
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25 -MODE=LOCAL
-OUTFMT=pc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10633202 @CGN 1.1 5180 @runat 14062005 141725 13099 -NCFU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp
-Q-/cgn2 1/USPTO spool p/US10633202/runat 14062005 141725 13099/app_query.fasta_1.903
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25 -MODE=LOCAL
-OUTFMT=pc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10633202 @CGN 1.1 5180 @runat 14062005 141725 13099 -NCFU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: gb_est1:
2: gb_est2:
3: gb_hc:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gssi:
9: gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2079.5	51.6	3245	3	BC036070 Homo sapi
2	2050	50.9	2169	9	AY405095 Mus muscu
3	2050	50.9	2912	3	AK014827 Mus muscu
4	1985.5	49.3	2615	3	AK016581 Mus muscu
5	1899.5	47.1	2871	3	AK029528 Mus muscu
6	1781.5	44.2	2551	3	AK029590 Mus muscu
7	1713.5	42.5	2490	3	AK029756 Mus muscu
8	1329.5	33.0	3881	3	BC047156 Mus muscu
9	1303	32.3	3511	3	AK077058 Mus muscu

10	1260	31.3	3128	3	BC047939	Homo sapi
11	1169.5	29.0	2164	3	AK029742	Mus muscu
12	1160.5	28.8	2492	3	AK076647	Mus muscu
13	1137.5	28.2	3174	3	AK016558	Mus muscu
14	1125	27.9	2546	3	AK076580	Mus muscu
15	1042	25.9	602	7	CV024284	Full
16	971	24.1	3108	3	AF326318	Homo sapi
17	948.5	23.5	2169	9	AY405093	Homo sapi
18	870.5	21.6	859	9	CR804792	GR0AAA26A
19	866.5	21.5	2547	3	AK016550	Mus muscu
20	829.5	20.6	2515	3	AK016574	Mus muscu
21	829.5	20.6	2547	3	AK029519	Mus muscu
22	827	20.5	2531	3	AK039384	Mus muscu
23	825	20.5	2976	3	AK048901	Mus muscu
24	804	20.0	804	9	CR815043	GR0AAA4BC
25	799.5	19.8	670	9	CE337023	CH230-509
26	798.5	19.8	788	8	BH363602	CH230-509
27	792	19.7	2235	3	AK005759	Mus muscu
28	779.5	19.3	2364	3	AK076633	Mus muscu
c	779	19.3	461	8	AQ060341	CIT-HSP-2
30	777	19.3	772	3	CR815044	GR0AAA4BC
31	776	19.3	2546	3	BC025386	Homo sapi
32	768.5	19.1	3036	3	AK089086	Mus muscu
33	767.5	19.0	2277	3	AK089169	Mus muscu
34	755.5	18.7	2891	3	AK046677	Mus muscu
c	743	18.4	631	1	AA400496	zu70505.s
36	736	18.3	2229	3	AK048367	Mus muscu
37	735.5	18.3	3433	3	AK029301	Mus muscu
38	734.5	18.2	3493	3	AK051780	Mus muscu
39	722.5	17.9	2661	9	AY399360	Homo sapi
40	719.5	17.9	3577	3	AK053825	Mus muscu
41	719	17.8	708	9	CR829312	GR0AAA60D
42	709.5	17.6	2664	9	AY399362	Mus muscu
43	696.5	17.3	1019	7	CN647602	ILLUMIGEN
44	695.5	17.3	2607	9	AY399361	Pan trogl
45	689.5	17.1	2250	9	AY418590	Homo sapi

ALIGNMENTS

RESULT 1	BC036070	3245 bp	mRNA	linear	HTC 20-SEP-2002
LOCUS	BC036070	Homo sapiens, clone IMAGE:5298677,	mRNA.		
DEFINITION	BC036070	BC036070.1	GI:23243314		
ACCESSION	BC036070	HTC.			
VERSION	BC036070.1	GI:23243314			
KEYWORDS	HTC.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens (human)				
REFERENCE	1	(bases 1 to 3245)			
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (31-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN) cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.				

Clone distribution: MGC clone distribution information can be found

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 16, 2005, 04:15:12 ; Search time 5730 Seconds
(without alignments)
786.506 Million cell updates/sec

Title: US-10-633-202-2
Perfect score: 4030
Sequence: 1 MAVGPELVHVRVTLILLWFG.....LFLCHVLPKRTKSKDEEG 726

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 6054689 seqs, 3103772919 residues

Total number of hits satisfying chosen parameters: 12109378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO spool_p/US10633202/runat_14062005_141727_13196/app query.fasta_1.903
-DB=Published Applications NA_QPWT-fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=25 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10633202 @CIGN 1.1 723 @runat_14062005_141727_13196
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:
22: /cgn2_6/ptodata/2/pubpna/US10J_NEW_PUB.seq:
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4030	100.0	2181	18	US-10-633-202-1	Sequence 1, Appli
2	4030	100.0	2578	21	US-10-956-157-1123	Sequence 1123, Ap
3	4030	100.0	3230	10	US-09-978-418-5	Sequence 5, Appli
4	4030	100.0	3230	21	US-10-485-231-5	Sequence 5, Appli
5	4004	99.4	2442	19	US-10-283-975A-397	Sequence 397, App
6	4004	99.4	2442	21	US-10-451-997-105	Sequence 105, App
7	2435	60.4	2169	14	US-10-265-125-1	Sequence 1, Appli
8	2435	60.4	2406	21	US-10-956-157-1122	Sequence 1122, Ap
9	2108.5	52.3	2958	21	US-10-956-157-1891	Sequence 7, Appli
10	2102.5	52.2	2301	18	US-10-664-456-7	Sequence 8, Appli
11	2102.5	52.2	2364	18	US-10-664-456-8	Sequence 9, Appli
12	2102.5	52.2	2463	18	US-10-664-456-9	Sequence 12, Appli
13	2102.5	52.2	2629	17	US-10-382-600-12	Sequence 20, Appli
14	2102.5	52.2	4007	17	US-10-382-600-17	Sequence 17, Appli
15	2102.5	52.2	5112	17	US-10-382-600-2	Sequence 2, Appli
16	2102.5	52.2	5165	17	US-10-382-600-2	Sequence 2, Appli
17	2102.5	52.2	6106	17	US-10-382-600-14	Sequence 14, Appli
18	2102.5	52.2	6159	17	US-10-382-600-1	Sequence 1, Appli
19	2102.5	52.2	6160	17	US-10-382-600-7	Sequence 7, Appli
20	2102.5	52.2	6160	17	US-10-382-600-11	Sequence 11, Appli
21	2102.5	52.2	6213	17	US-10-382-600-18	Sequence 18, Appli
22	2102.5	52.2	6213	17	US-10-382-600-21	Sequence 21, Appli
23	2102.5	52.2	6214	17	US-10-382-600-4	Sequence 4, Appli
24	2102.5	52.2	6267	17	US-10-382-600-15	Sequence 15, Appli
25	1654	41.0	1653	9	US-09-792-200B-11	Sequence 11, Appli
26	1387.5	34.4	2373	16	US-10-156-028-1	Sequence 1, Appli
27	1387.5	34.4	2695	14	US-10-028-072-203	Sequence 203, App
28	1387.5	34.4	2695	14	US-10-140-808-203	Sequence 203, App
29	1387.5	34.4	2695	14	US-10-121-049-203	Sequence 203, App
30	1387.5	34.4	2695	14	US-10-123-904-203	Sequence 203, App
31	1387.5	34.4	2695	14	US-10-140-470-203	Sequence 203, App
32	1387.5	34.4	2695	14	US-10-175-746-203	Sequence 203, App
33	1387.5	34.4	2695	14	US-10-176-918-203	Sequence 203, App
34	1387.5	34.4	2695	14	US-10-176-921-203	Sequence 203, App
35	1387.5	34.4	2695	14	US-10-137-865-203	Sequence 203, App
36	1387.5	34.4	2695	14	US-10-140-474-203	Sequence 203, App
37	1387.5	34.4	2695	14	US-10-142-431-203	Sequence 203, App
38	1387.5	34.4	2695	14	US-10-143-114-203	Sequence 203, App
39	1387.5	34.4	2695	14	US-10-142-419-203	Sequence 203, App
40	1387.5	34.4	2695	14	US-10-123-262-203	Sequence 203, App
41	1387.5	34.4	2695	14	US-10-142-423-203	Sequence 203, App
42	1387.5	34.4	2695	14	US-10-121-050-203	Sequence 203, App
43	1387.5	34.4	2695	14	US-10-141-755-203	Sequence 203, App
44	1387.5	34.4	2695	14	US-10-143-032-203	Sequence 203, App
45	1387.5	34.4	2695	14	US-10-123-108-203	Sequence 203, App

ALIGNMENTS

RESULT 1
US-10-633-202-1
; Sequence 1, Application US/10633202
; Publication No. US2004005314A1
; GENERAL INFORMATION:
; APPLICANT: CERRETTI, Douglas P.
; TITLE OF INVENTION: SVPH1-26 DNA AND POLYPEPTIDES
; FILE REFERENCE: 03260.0036-00304
; CURRENT APPLICATION NUMBER: US/10/633,202
; PRIOR FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: US/09/561,779
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/063,571
; PRIOR FILING DATE: 1997-10-30
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 16, 2005, 00:23:41 ; Search time 432 Seconds
(without alignments)
2749.854 Million cell updates/sec

Title: US-10-633-202-2

Perfect score: 4030

Sequence: 1 MAVGBPLVHRVTLILLWFG.....LFLHLVLFKRTKSKDEBEG 726

Scoring table:

BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlp
-Q/cgn2_1/USPTO_spool_p/US10633202/runat 14062005 141726 13120/app query.fasta_1.903
-DB=issued Patents NA -QFMT=fastcap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10633202 @CEN 1.1 105 @runat 14062005 141726 13120 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	4030	100.0	35837	4	US-09-949-016-15232
2	2435	60.4	2169	4	US-09-617-145-1
3	2435	60.4	3557	4	US-09-949-016-547
4	2376	59.0	7557	4	US-09-949-016-12289
5	2108.5	52.3	2959	4	US-09-949-016-3341
6	2108.5	52.3	2959	4	US-09-949-016-3342
7	2108.5	52.3	2959	4	US-09-949-016-3343
8	2108.5	52.3	50850	4	US-09-949-016-15083
9	2108.5	52.3	50850	4	US-09-949-016-15084
10	2108.5	52.3	50850	4	US-09-949-016-15085
11	1387.5	34.4	2373	4	US-09-949-016-2298
12	1387.5	34.4	2373	4	US-09-949-016-2299

13	1387.5	34.4	6373	4	US-09-949-016-12830	Sequence 12830, A
14	1387.5	34.4	6373	4	US-09-949-016-12888	Sequence 12888, A
15	1387.5	34.4	6374	4	US-09-949-016-14040	Sequence 14040, A
16	1387.5	34.4	6374	4	US-09-949-016-14041	Sequence 14041, A
17	1386.5	34.4	2346	4	US-09-949-016-1088	Sequence 1088, Ap
18	1385.5	34.4	2373	4	US-09-949-016-1146	Sequence 1146, Ap
19	1317	32.7	2207	4	US-09-949-016-4593	Sequence 4593, Ap
20	1175.5	29.2	3758	4	US-09-976-594-491	Sequence 491, App
21	1009.5	25.0	5062	4	US-09-949-016-499	Sequence 499, App
22	980.5	24.3	4471	4	US-09-949-016-4910	Sequence 4910, Ap
23	955.5	23.7	2976	4	US-09-949-016-5358	Sequence 5358, Ap
24	955.5	23.7	3261	4	US-09-949-016-2391	Sequence 2391, Ap
25	924	22.9	3468	3	US-09-632-098-3	Sequence 3, Appli
26	917.5	22.8	3468	4	US-10-177-308-3	Sequence 3, Appli
27	917.5	22.8	3582	4	US-09-548-797B-3	Sequence 3, Appli
28	911	22.6	2650	5	PCT-US95-07295-7	Sequence 7, Appli
29	911	22.6	2650	5	PCT-US95-07295-7	Sequence 7, Appli
30	885	22.0	3390	4	US-09-548-797B-2	Sequence 2, Appli
31	884.5	21.9	2556	4	US-10-000-489-69	Sequence 69, Appl
32	882.5	21.9	2325	4	US-09-786-256C-16	Sequence 16, Appl
33	882.5	21.9	2828	4	US-09-786-256C-31	Sequence 31, Appl
34	878	21.8	2373	1	US-08-264-101-1	Sequence 1, Appli
35	878	21.8	2373	2	US-08-765-243-1	Sequence 1, Appli
36	878	21.8	2373	5	PCT-US95-07295-1	Sequence 5, Appli
37	876.5	21.7	2553	2	US-08-765-243-5	Sequence 5, Appli
38	876.5	21.7	2553	3	PCT-US95-07295-5	Sequence 5, Appli
39	855.5	21.2	3431	5	US-09-632-098-1	Sequence 1, Appli
40	855.5	21.2	3431	4	US-10-177-308-1	Sequence 1, Appli
41	839	20.8	2648	2	US-08-836-443-1	Sequence 1, Appli
42	826	20.5	3994	4	US-09-738-945-7	Sequence 7, Appli
43	815.5	20.2	3271	4	US-09-548-797B-1	Sequence 1, Appli
44	808	20.0	1851	3	US-09-608-790-2	Sequence 2, Appli
45	799.5	19.8	2251	2	US-08-836-443-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-949-016-15232/c
; Sequence 15232, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15232
; LENGTH: 35837
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15232

Alignment Scores:
Pred. No.: 0 Length: 35837
Score: 4030.00 Matches: 726
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-633-202-2 (1-726) x US-09-949-016-15232 (1-35837)

QY 1 MetAlaValGlyGluProLeuValHisIleArgValThrLeuLeuLeuTrpPheGly 20

Db 2336 ATGCGAGTGGGTGAGCCCTGGTGCACATCAGGGTCACTCTCTGCTGCTGCTGGG 2277
Qy 21 MetPheLeuSerIleSerGlyHisSerGlnAlaArgProSerGlnTyrPheThrSerPro 40
Db 2276 ATGTTTTTGTCTATTTCTGGCCACTCTCAGGCCAGGCCCTCCAGTAGTATTTTCACCTTCCCA 2217
Qy 41 GluValValIleProLeuLysValIleSerArgGlyArgGlyAlaLysAlaProGlyTyr 60
Db 2216 GAAGTGGTGATCCCTTTGAAGGTGATCAGCAGGGGAGAGGTGCAGAGGCTCTCGNATGG 2157
Qy 61 LeuSerTyrSerLeuArgPheGlyGlyGlnArgTyrIleValHisMetArgValAsnLys 80
Db 2156 CTCTCCTATAGCCTGCGGTTGGGGGACAGAGATACATTGTGCACATGAGGGTAAATAAG 2097
Qy 81 LeuLeuPheAlaAlaHisLeuProValPheThrTyrThrGluGlnHisAlaLeuLeuGln 100
Db 2096 CTGTGTGTGTGTCGACACCTTCCTGTGTTCACCTACACAGAGCAGCATGCCCTGCTCCAG 2037
Qy 101 AspGlnProPheIleGlnAspAspCysTyrTyrHisGlyTyrValGluGlyValProGlu 120
Db 2036 GATCAGCCCTTCATCCAGGATGACTGCTACTACCATGGTTATGTGGAGGGGTCCTTGAG 1977
Qy 121 SerLeuValAlaLeuSerThrCysSerGlyGlyPheLeuGlyMetLeuGlnIleAsnAsp 140
Db 1976 TCCTTGGTGGCTTAGTACCTGTCTGGGGCTTCTTGGAAATGCTACAGATAAATGAC 1917
Qy 141 LeuValTyrGluIleLysProIleSerValSerAlaThrPheGluHisLeuValTyrLys 160
Db 1916 CTTCTTTATGAAATCAAGCCCAATTAGTGTTCGCCACATTTGAAACCTAGTATATAAG 1857
Qy 161 IleAspSerAspThrGlnPheProMetArgCysGlyLeuThrGluGluLysIle 180
Db 1856 ATAGCAGTGATGATACACAGTTCCACCTATGAGATGTGGTTAAACAGAGAAATA 1797
Qy 181 AlaHisGlnMetGluLeuGlnLeuSerTyrAsnPheThrLeuLysGlnSerSerPheVal 200
Db 1796 GCACACCAGATGGAGTTGCAATTTGTCATATAAATTTCACTCTGAAGCAAAAGTTCTTTGTG 1737
Qy 201 GlyTrpTrpThrHisGlnArgPheValGluLeuValValValValAspAsnIleArgTyr 220
Db 1736 GGCTGGTGGACCCATCAGCGGTTGTTGAGCTGGTAGTGTGGGATAATATATAGATAT 1677
Qy 221 LeuPheSerGlnSerAsnAlaThrThrValGlnHisGluValPheAsnValAlaAsnIle 240
Db 1676 CTTTTCTCTCAAAGTATGCAACACAGTGCAGATGAGATATTAACTGTGTCAATATA 1617
Qy 241 ValAspSerPheTyrHisProLeuGluValAspValIleLeuThrGlyIleAspIleTyr 260
Db 1616 GTGGATTCCTCTATCATCTCTTGGAGGTTGATGTAATTTTGAATTTGATATATAGG 1557
Qy 261 ThrAlaSerAsnProLeuProThrSerGlyAspLeuAspAsnValLeuGluAspPheSer 280
Db 1556 ACTGCATCAAAATCCACTTCCTACAGTGGAGACCTAGATAATGTTTTAGAGGACTTTTCT 1497
Qy 281 IleTrpLysAsnTyrAsnLeuAsnAsnArgLeuGlnHisAspValAlaHisLeuPheIle 300
Db 1496 ATTTGGAGAAATTAATACCTTAATAATCGACTACACATGATGTGTGCACATCTTTTCATA 1437
Qy 301 LysAspThrGlnGlyMetLysLeuGlyValAlaTyrValLysGlyIleCysGlnAsnPro 320
Db 1436 AAAGACACACAAGGCATGAAGCTTGGTGTGGCTTATGTTAAAGGAATATGCCAGAACTCT 1377
Qy 321 PheAsnThrGlyValAspValPheGluAspAsnArgLeuValValPheAlaIleThrLeu 340
Db 1376 TTTAAATCTAGGTTGATGTTTTTGAAGACAACAGGTGTGGCTGTTTTGCAATTTACTTTG 1317
Qy 341 GlyHisGluLeuGlyHisAsnLeuGlyMetGlnHisAspThrGlnTrpCysValCysGlu 360
Db 1316 GGCCACGAGCTTGGTCATATTTGGGTATGCACATGACACCCAGTGGTGTGTGTGCGAG 1257
Qy 361 LeuGlnTrpCysIleMetHisAlaTyrArgLysValThrThrLysPheSerAsnCysSer 380

Db 1256 CTACAGTGGTGCATATATGCATGCTATAGAAAGGTGACAACATAAATTTAGCAACTCAGT 1197
Qy 381 TyrAlaGlnTyrTrpAspSerThrIleSerSerGlyLeuCysIleGlnProProProTyr 400
Db 1196 TATGCCAATATTTGGGACAGTACTATCAGTAGTGGATTATGATTATTCACCGCCTCCATAT 1137
Qy 401 ProGlyAsnIlePheArgLeuLysTyrCysGlyAsnLeuValValGluGluGlyGlu 420
Db 1136 CCAGGGAATATATTTAGACTGAAGTACTGTGGGAATCTAGTGGTGAAGAGGGAGGAA 1077
Qy 421 CysAspCysGlyThrIleArgGlnCysAlaLysAspProCysCysLeuLeuAsnCysThr 440
Db 1076 TGTGACTGTGGACCACTACGCGAGTGTGCAAAAGATCCCTGTGTGTGTTAAACTGTACT 1017
Qy 441 LeuHisProGluValAlaAlaCysAlaPheGlyIleCysCysLysAspCysLysPheLeuPro 460
Db 1016 CTACATCTCGGGCTCCTTGTCTTTGGAAATATGTTGCAAGACTGCAAAATTTCTGCCA 957
Qy 461 SerGlyThrLeuCysArgGlnGlnValGlyGluCysAspLeuProGluTrpCysAsnGly 480
Db 956 TCAGGAACCTTATGTAGACAACAAGTTGGTGAATGTGACCTTCCAGAGTGGTGCATGGG 897
Qy 481 ThrSerHisGlnCysProAspAspValTyrValGlnAspGlyIleSerCysAsnValAsn 500
Db 896 ACATCCCATCAATGCCAGATGATGTATGTGACGACGGGATCTCTGTATATGTGAAT 837
Qy 501 AlaPheCysTyrGluLysThrCysAsnAsnHisAspIleGlnCysLysGluIlePheGly 520
Db 836 GCCTTCTGCTATGAAAGACGTGTAATAACCATGATATACAATGTAAAGAGATTTTGGC 777
Qy 521 GlnAspAlaArgSerAlaSerGlnSerCysTyrGlnGluIleAsnThrGlnGlyAsnArg 540
Db 776 CRAAGTCAAGGAGTGCATCTCAGAGTTGCTACCAAGAAATCAACACCAAGGAAACCGT 717
Qy 541 PheGlyHisCysGlyIleValGlyThrThrTyrValLysCysTrpThrProaspIleMet 560
Db 716 TTGCGTCACTGTGTTATTTAGGCACACATATGTAATAATGTTGGACCCCTGTATATCATG 657
Qy 561 CysGlyArgValGlnCysGluAsnValGlyValIleProAsnLeuIleGluHisSerThr 580
Db 656 TGTGGAGGGTTTCAGTGTGAAAATGTGGAGTAAATTCCCAATCTGATAGAGCATCTTACA 597
Qy 581 ValGlnGlnPheHisLeuAsnAspThrThrCysTyrGlyThrAspTyrHisLeuGlyMet 600
Db 596 GTGCACAGTTTCACTCACTCAATGACACCACTTGTGCGGCACTGATTTATCATTTAGGGATG 537
Qy 601 AlaIleProAspIleGlyGluValLysAspGlyThrValCysGlyProGluLysIleCys 620
Db 536 GCTATACCTGATATTTGGTGGGTGAAAGATGGCACAGTATGTGGTCCAGAAAAGATCTGC 477
Qy 621 IleArgLysLysCysAlaSerMetValHisLeuSerGlnAlaCysGlnProLysThrCys 640
Db 476 ATCCGTAAAGAGTGTGCCAGTATGGTTTCATGTGTCAAGCCCTGTGAGCTTAGACCTGC 417
Qy 641 AsnMetArgGlyIleCysAsnAsnLysGlnHisCysHisCysAsnHisGluTrpAlaPro 660
Db 416 AACATGAGGGGAATCTGCACAACAACAACCACTGTCTGCAACCATGAAATGGGCACCC 357
Qy 661 ProTyrCysLysAspLysGlyTyrGlyLysArgAlaAspSerGlyProProProLysAsn 680
Db 356 CCATACCTGCAAGGACCAAGGCTATGGAGGTAGTGTCTGATAGTGGCCACCTCTCTAAGAAC 297
Qy 681 AsnMetGluGlyLeuAsnValMetGlyLysLeuArgTyrLeuSerLeuLeuCysLeuLeu 700
Db 296 AACATGGAAGGATTAATGTGATGGGAAAGTTGCGTTTACTCTGTCACTATTTGCGCTTCT 237
Qy 701 ProLeuValAlaPheLeuLeuPheCysLeuHisValLeuPheLysLysArgThrLysSer 720
Db 236 CCTTTGTTGCTTTTATTTATTTTGTCTTACATGTCTTTTAAAGAAACGACACAAAAGT 177
Qy 721 LysGluAspGluGly 726
Db 176 AAAGAAGATGAAGAAGGA 159

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 20:14:51 ; Search time 110 Seconds
(without alignments)
2530.007 Million cell updates/sec

Title: US-10-633-202-2

Perfect score: 4030

Sequence: 1 MAVGSLVHVRITLLLMFG.....LFLHLVFKRTSKEDDEG 726

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1710399 seqs, 383334425 residues

Total number of hits satisfying chosen parameters: 1710399

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4030	100.0	726	15 US-10-633-202-2	Sequence 2, Appli
2	4030	100.0	776	17 US-09-978-418-6	Sequence 6, Appli
3	4030	100.0	776	17 US-10-485-231-6	Sequence 6, Appli
4	4004	99.4	726	14 US-10-156-028-3	Sequence 3, Appli
5	4004	99.4	726	17 US-10-491-997-106	Sequence 106, App
6	2435	60.4	722	14 US-10-265-125-2	Sequence 2, Appli
7	2435	60.4	722	15 US-10-074-978A-289	Sequence 289, App
8	2427	60.2	722	15 US-10-074-978A-288	Sequence 288, App
9	2102.5	52.2	766	15 US-10-664-456-12	Sequence 12, Appl
10	2102.5	52.2	787	15 US-10-664-456-13	Sequence 13, Appl
11	2102.5	52.2	820	15 US-10-664-456-14	Sequence 14, Appl

12	2102.5	52.2	820	16	US-10-408-765A-1252	Sequence 1252, Ap
13	2050	50.9	729	15	US-10-074-978A-287	Sequence 287, App
14	1654	41.0	446	9	US-09-792-200B-12	Sequence 12, Appl
15	1646	40.8	535	15	US-10-633-202-3	Sequence 3, Appli
16	1387.5	34.4	790	14	US-10-028-072-204	Sequence 204, App
17	1387.5	34.4	790	14	US-10-140-808-204	Sequence 204, App
18	1387.5	34.4	790	14	US-10-121-049-204	Sequence 204, App
19	1387.5	34.4	790	14	US-10-123-904-204	Sequence 204, App
20	1387.5	34.4	790	14	US-10-140-470-204	Sequence 204, App
21	1387.5	34.4	790	14	US-10-175-746-204	Sequence 204, App
22	1387.5	34.4	790	14	US-10-176-918-204	Sequence 204, App
23	1387.5	34.4	790	14	US-10-176-921-204	Sequence 204, App
24	1387.5	34.4	790	14	US-10-137-855-204	Sequence 204, App
25	1387.5	34.4	790	14	US-10-140-474-204	Sequence 204, App
26	1387.5	34.4	790	14	US-10-142-431-204	Sequence 204, App
27	1387.5	34.4	790	14	US-10-143-114-204	Sequence 204, App
28	1387.5	34.4	790	14	US-10-142-419-204	Sequence 204, App
29	1387.5	34.4	790	14	US-10-123-262-204	Sequence 204, App
30	1387.5	34.4	790	14	US-10-142-423-204	Sequence 204, App
31	1387.5	34.4	790	14	US-10-121-050-204	Sequence 204, App
32	1387.5	34.4	790	14	US-10-141-755-204	Sequence 204, App
33	1387.5	34.4	790	14	US-10-143-032-204	Sequence 204, App
34	1387.5	34.4	790	14	US-10-123-108-204	Sequence 204, App
35	1387.5	34.4	790	14	US-10-123-236-204	Sequence 204, App
36	1387.5	34.4	790	14	US-10-123-261-204	Sequence 204, App
37	1387.5	34.4	790	14	US-10-140-921-204	Sequence 204, App
38	1387.5	34.4	790	14	US-10-140-928-204	Sequence 204, App
39	1387.5	34.4	790	14	US-10-121-045-204	Sequence 204, App
40	1387.5	34.4	790	14	US-10-123-292-204	Sequence 204, App
41	1387.5	34.4	790	14	US-10-123-903-204	Sequence 204, App
42	1387.5	34.4	790	14	US-10-124-819-204	Sequence 204, App
43	1387.5	34.4	790	14	US-10-124-822-204	Sequence 204, App
44	1387.5	34.4	790	14	US-10-140-925-204	Sequence 204, App
45	1387.5	34.4	790	14	US-10-160-498-204	Sequence 204, App

ALIGNMENTS

RESULT 1

US-10-633-202-2
; Sequence 2, Application US/10633202
; Publication NO. US20040053141A1
; GENERAL INFORMATION:
; APPLICANT: CERRETTI, Douglas P.
; TITLE OF INVENTION: SVPH1-26 DNA AND POLYPEPTIDES
; FILE REFERENCE: 03260.0036-00304
; CURRENT FILING DATE: 2003-07-29
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US/09/561,779
; PRIOR FILING DATE: 1997-10-30
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-633-202-2

Query Match 100.0%; Score 4030; DB 15; Length 726;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAVGSLVHVRITLLLMFGMFLSISGHSQARPSQYFTSPVVIPLKVISRGKAKPGW	60
Db	1	MAVGSLVHVRITLLLMFGMFLSISGHSQARPSQYFTSPVVIPLKVISRGKAKPGW	60
Qy	61	LSVSLRFGGQRIYVHMRVKNLLFAAHLPVFTTTEQHALLQDQPFQIDDCYHYHVGVEPE	120
Db	61	LSVSLRFGGQRIYVHMRVKNLLFAAHLPVFTTTEQHALLQDQPFQIDDCYHYHVGVEPE	120

121 SLVALSTCSGGFLGMLQINDLVYKIPISVSATFEHLVYKIDSDDTQFPMPRCGLTEEKI 180
121 SLVALSTCSGGFLGMLQINDLVYKIPISVSATFEHLVYKIDSDDTQFPMPRCGLTEEKI 180
181 AHQMEIQLSNTFLTKOSSFVGMWTHQRFVELVVDNIRYLFQSQSNATTVOHEVFVNVI 240
181 AHQMEIQLSNTFLTKOSSFVGMWTHQRFVELVVDNIRYLFQSQSNATTVOHEVFVNVI 240
241 VDSFYHPLVDVILTGIDIWTSNPLPTSGDLNVLDFSIWKYNNLNRLQHDVAHLFI 300
241 VDSFYHPLVDVILTGIDIWTSNPLPTSGDLNVLDFSIWKYNNLNRLQHDVAHLFI 300
301 KDTQGMKLGVAAYVKGICQNPFTGVDVFEDNRLVFAITLGHGELHNLGQHDTQMCVCE 360
301 KDTQGMKLGVAAYVKGICQNPFTGVDVFEDNRLVFAITLGHGELHNLGQHDTQMCVCE 360
361 LQWCIHAYRKVTKFSNCSYAQYWDSTISSGLCIQPPYPGNI FRLKYCGNLVVEEGE 420
361 LQWCIHAYRKVTKFSNCSYAQYWDSTISSGLCIQPPYPGNI FRLKYCGNLVVEEGE 420
421 CDCGTTIRQCAKDPCCLLNCTLHPGAACAFGICCKDCKFLPSGTLCRQQVGECDLPEWCNG 480
421 CDCGTTIRQCAKDPCCLLNCTLHPGAACAFGICCKDCKFLPSGTLCRQQVGECDLPEWCNG 480
481 TSHQCDDVVYQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQSCYQBIINTQGNR 540
481 TSHQCDDVVYQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQSCYQBIINTQGNR 540
541 FGHCGIVGTTYVKCWTPTDIMCGRVQCNVGI PNLIHSTVQOFLHNDTTCWGTDYHLGM 600
541 FGHCGIVGTTYVKCWTPTDIMCGRVQCNVGI PNLIHSTVQOFLHNDTTCWGTDYHLGM 600
601 AIPDGEVKDGTVCGEPEKICIRKKCASMVHLSQACOPKTCNNMRGICNNKQHCNHEWAP 660
601 AIPDGEVKDGTVCGEPEKICIRKKCASMVHLSQACOPKTCNNMRGICNNKQHCNHEWAP 660
661 PYCKDKGYGGSADSGPPPKNNMEGLNVMGKRLYSLCLLPLVAFLLCFLHVLFKKRTKS 720
661 PYCKDKGYGGSADSGPPPKNNMEGLNVMGKRLYSLCLLPLVAFLLCFLHVLFKKRTKS 720
721 KEDEEG 726
721 KEDEEG 726

RESULT 2
US-09-978-418-6
; Sequence 6, Application US/09978418
; Publication No. US20030118997A1
; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephan
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 142.US5.REG
; CURRENT APPLICATION NUMBER: US/09/978,418
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: 60/311,305
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/314,734
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/318,204
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/326,470
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: JPatent
; SEQ ID NO 6
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-978-418-6

Query Match 100.0%; Score 4030; DB 10; Length 776;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAVGEPLVHVRVTLTLLWFGMELSI SGHSQARPSQYFTSPVVIPLKVISRGRGAKAPGW 60
DB 51 MAVGEPLVHVRVTLTLLWFGMELSI SGHSQARPSQYFTSPVVIPLKVISRGRGAKAPGW 110
QY 61 LYSLSRFGGORYIVHMRVKNKLLFAAHLPVFTTTEQHALLQDQDQFIQDDCYHYHGYVGEVPE 120
DB 111 LYSLSRFGGORYIVHMRVKNKLLFAAHLPVFTTTEQHALLQDQDQFIQDDCYHYHGYVGEVPE 170
QY 121 SLVALSTCSGGFLGMLQINDLVYKIPISVSATFEHLVYKIDSDDTQFPMPRCGLTEEKI 180
DB 171 SLVALSTCSGGFLGMLQINDLVYKIPISVSATFEHLVYKIDSDDTQFPMPRCGLTEEKI 230
QY 181 AHQMEIQLSNTFLTKOSSFVGMWTHQRFVELVVDNIRYLFQSQSNATTVOHEVFVNVI 240
DB 231 AHQMEIQLSNTFLTKOSSFVGMWTHQRFVELVVDNIRYLFQSQSNATTVOHEVFVNVI 290
QY 241 VDSFYHPLVDVILTGIDIWTSNPLPTSGDLNVLDFSIWKYNNLNRLQHDVAHLFI 300
DB 291 VDSFYHPLVDVILTGIDIWTSNPLPTSGDLNVLDFSIWKYNNLNRLQHDVAHLFI 350
QY 301 KDTQGMKLGVAAYVKGICQNPFTGVDVFEDNRLVFAITLGHGELHNLGQHDTQMCVCE 360
DB 351 KDTQGMKLGVAAYVKGICQNPFTGVDVFEDNRLVFAITLGHGELHNLGQHDTQMCVCE 410
QY 361 LQWCIHAYRKVTKFSNCSYAQYWDSTISSGLCIQPPYPGNI FRLKYCGNLVVEEGE 420
DB 411 LQWCIHAYRKVTKFSNCSYAQYWDSTISSGLCIQPPYPGNI FRLKYCGNLVVEEGE 470
QY 421 CDCGTTIRQCAKDPCCLLNCTLHPGAACAFGICCKDCKFLPSGTLCRQQVGECDLPEWCNG 480
DB 471 CDCGTTIRQCAKDPCCLLNCTLHPGAACAFGICCKDCKFLPSGTLCRQQVGECDLPEWCNG 530
QY 481 TSHQCDDVVYQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQSCYQBIINTQGNR 540
DB 531 TSHQCDDVVYQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQSCYQBIINTQGNR 590
QY 541 FGHCGIVGTTYVKCWTPTDIMCGRVQCNVGI PNLIHSTVQOFLHNDTTCWGTDYHLGM 600
DB 591 FGHCGIVGTTYVKCWTPTDIMCGRVQCNVGI PNLIHSTVQOFLHNDTTCWGTDYHLGM 650
QY 601 AIPDGEVKDGTVCGEPEKICIRKKCASMVHLSQACOPKTCNNMRGICNNKQHCNHEWAP 660
DB 651 AIPDGEVKDGTVCGEPEKICIRKKCASMVHLSQACOPKTCNNMRGICNNKQHCNHEWAP 710
QY 661 PYCKDKGYGGSADSGPPPKNNMEGLNVMGKRLYSLCLLPLVAFLLCFLHVLFKKRTKS 720
DB 711 PYCKDKGYGGSADSGPPPKNNMEGLNVMGKRLYSLCLLPLVAFLLCFLHVLFKKRTKS 770
QY 721 KEDEEG 726
DB 771 KEDEEG 776

RESULT 3
US-10-485-231-6
; Sequence 6, Application US/10485231
; Publication No. US20050119171A1
; GENERAL INFORMATION:
; APPLICANT: GENSET
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 142 WO1
; CURRENT APPLICATION NUMBER: US/10/485,231
; CURRENT FILING DATE: 2004-01-29
; PRIOR APPLICATION NUMBER: 60/311,305
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/314,734
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/318,204
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/326,470

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 19:52:20 ; Search time 56 Seconds
(without alignments)
967.772 Million cell updates/sec

Title: US-10-633-202-2
Perfect score: 4030
Sequence: 1 MAVGSLVHVRVTLMLWFG.....LFCLHVLKPKTKSKEDDEG 726

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/prodata/1/1aa/PTUS_COMB.pep.*
6: /cgn2_6/prodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2435	60.4	722	US-09-617-145-2	Sequence 2, Appli
2	2435	60.4	722	US-09-949-016-6418	Sequence 6418, Ap
3	2108.5	52.3	826	US-09-949-016-9212	Sequence 9212, Ap
4	2108.5	52.3	826	US-09-949-016-9213	Sequence 9213, Ap
5	2108.5	52.3	826	US-09-949-016-9214	Sequence 9214, Ap
6	1387.5	34.4	790	US-09-949-016-8169	Sequence 8169, Ap
7	1387.5	34.4	790	US-09-949-016-8170	Sequence 8170, Ap
8	1386.5	34.4	781	US-09-949-016-6959	Sequence 6959, Ap
9	1385.5	34.4	790	US-09-949-016-7017	Sequence 7017, Ap
10	1314.5	32.6	705	US-09-949-016-10464	Sequence 10464, A
11	1009.5	25.0	909	US-09-949-016-6370	Sequence 6370, Ap
12	955.5	23.7	953	US-09-949-016-8262	Sequence 8262, Ap
13	955.5	23.7	968	US-09-949-016-11229	Sequence 11229, A
14	924	22.9	812	US-09-632-098-4	Sequence 4, Appli
15	924	22.9	812	US-10-177-308-4	Sequence 4, Appli
16	917.5	22.8	849	US-09-548-797B-6	Sequence 6, Appli
17	910	22.6	734	US-08-765-243-8	Sequence 8, Appli
18	910	22.6	734	PCT-US95-07295-8	Sequence 8, Appli
19	885	22.0	787	US-09-348-797B-5	Sequence 5, Appli
20	882.5	21.9	775	US-09-786-256C-15	Sequence 15, Appl
21	882.5	21.9	775	US-09-786-256C-32	Sequence 32, Appl
22	879.5	21.8	787	US-10-000-489-70	Sequence 70, Appl
23	878	21.8	651	US-08-264-101-2	Sequence 2, Appli
24	878	21.8	651	US-08-765-243-2	Sequence 2, Appli
25	878	21.8	651	PCT-US95-07295-2	Sequence 2, Appli
26	873.5	21.7	735	US-08-765-243-6	Sequence 6, Appli
27	873.5	21.7	735	PCT-US95-07295-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-09-617-145-2
; Sequence 2, Application US/09617145
; Patent No. 6485956
; GENERAL INFORMATION:
; APPLICANT: Cerretti, Douglas P.
; TITLE OF INVENTION: SVPHI-8 DNA and Polypeptides
; FILE REFERENCE: 03260.0050-00304
; CURRENT APPLICATION NUMBER: US/09/617,145
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/071,505
; PRIOR FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-617-145-2

Query Match 60.4%; Score 2435; DB 4; Length 722;

Best Local Similarity 61.6%; Pred. No. 2e+207;

Matches 444; Conservative 101; Mismatches 172; Indels 4; Gaps 4;

Qy	1	MAVGEPLVHVRVTLMLWFGFLSISGHSQAPSPQYFTSPVVIPLKVISRGKAPGM 60	Sequence 2, Appli
Db	1	MAVDGTLVIRVTLMLWFGFLSISGYSQAPSPQYFTSPVVIPLKVISRGKAPGM 60	Sequence 6418, Ap
Qy	61	LSYSLRFGGQRYIVHMYNKLFAAHLPVFTVTEQHALLQDQFFIQQDCYHYHGVGVE 120	Sequence 9212, Ap
Db	61	LSYSLRFGGQRYIVHMYNKLFAAHLPVFTVTEQHALLQDQFFIQQDCYHYHGVGVE 120	Sequence 9213, Ap
Qy	121	SLVALSTCGGFLGMLQINDLVYEIKPISVSATFEHLVYKIDSDTQPPMRCGLTEEKI 180	Sequence 8169, Ap
Db	121	SLVVSACFGGFLGMLQINDLVYEIKPISVSATFEHLVYKIDSDTQPPMRCGLTEEKI 180	Sequence 8170, Ap
Qy	181	A-HOMELQSLNFTYKQSSFGVQWTHQRPVELVVDVIRYLFPSQSNATTQHEVPNVN 239	Sequence 6959, Ap
Db	181	ARQQLFEFEAEASALEPKSAGDWTHAWFLELVVVDVIRYLFPSQSNATTQHEVPNVN 240	Sequence 7017, Ap
Qy	240	IVDSFVHPLVDVITLTDITWASPLPTSGDLNVDLDFSTWKNYNLNRLQHDVAHLF 299	Sequence 10464, A
Db	241	IVDSWYQQLGTVTIILIGIEIMNQGVFPMT-SIEQLNDFSQWKOISL-SQLQDAAHNF 298	Sequence 6370, Ap
Qy	300	IKDTCMKLGVAYVXGICQNPENTGVDFEDNRLVVFATLGHGLHNLGMQHDQWCV 359	Sequence 8262, Ap
Db	299	IKNSLISILGLAYVAGICRPDIDCGVDNFQGDWLSLFANTVAHLGHTLGMQHDDEFCFC 358	Sequence 11229, A
Qy	360	ELQWCMIMAYRKVTTKFNCSVAQWDSITSSGLCIQPPYPGNIFRLKYCNLVVESE 419	Sequence 4, Appli

Db 359 GERCIMNTRVPAEFTNCSYADFMKTTTLNQSGSLNPPRLGEIFMLKRCGNGVVEREE 418
Qy 420 EDCGTTIRQCAKDPCCLLNCTLHPGAACAFGICCKCKFLPSGTLCRQQVGECDLPEWCN 479
Db 419 QDCGVSQCEQDACCLLNCTLRGAACAFGLCKCKFMPSGELCHQEVNECDLPEWCN 478
Qy 480 GTSHQCPDDVYVQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQSCYQEIINTQGN 539
Db 479 GTSHQCPEDRYVQDGIPCSDSAYCYQRCNNHDOHCREIFGDKAKSASQNCYKEINSQGN 538
Qy 540 RFHCGIGVTYVYKCTPDMCGRVQCEVGVIPNLIEHSTVQOQFHLNDTTCWGTDYHLG 599
Db 539 RFHCGINGITYLKCHISDVFCGRVQCEVNRDIPLLQDHFTLQHTHINGVTCWGDYHLR 598
Qy 600 MAIPDIEGVKDGTVCGPEKICIRKCKASVMHLSQACQPKTCNMRGICNNKHQCHCHNEWA 659
Db 599 MNISDIEGVKDGTVCGPGKICIKKCVLSVLSHVCLPETCNMKGICNNKHCHCHGCGWS 658
Qy 660 PPYCKDKYGGSDSGPPKNNMEGLNMGKRLYLSLLCLLPLVAFLLFCLHVLFFKRTK 719
Db 659 PPYQHRGYGGSIDSGPASAKRGVFLPLI-VIPSLSVLTFTVGLMYLRLQCSGPKETK 717
Qy 720 S 720
Db 718 A 718
RESULT 2
US-09-949-016-6418
; Sequence 6418, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6418
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6418
Query Match 60.4%; Score 2435; DB 4; Length 722;
Best Local Similarity 61.6%; Pred. No. 2e-207;
Matches 444; Conservative 101; Mismatches 172; Indels 4; Gaps 4;
Qy 1 MAYGEPIVHVRVTLILLWGMFLSISGHSQARPSQFTSPSEVVIPLKVISRGRGAKAPGW 60
Db 1 MAVDGLTVIRVTLILLWGMFLSISGYSQAGFSQHTSPSEVVIPLKVISRGRGAKAPGW 60
Qy 61 LSYSLRPGGQRYIVHMRVKNLLFAAHLPVFTYTEQHALLODQPFIODDCYHYGVGEVPE 120
Db 61 LSYSLRPGGQRYVHMRVKNLLVSRHLPVFTYTERALLEQDLFPDDCYHYGVGEAPE 120
Qy 121 SLVALSCSGGFGMLQINDLVVEIKPISVATFEHLVYKIDSDDDTQFPFMRGCLTEEKI 180
Db 121 SLVVFSACGFGGVLKISGLTYEIEPIRSATFEHLVYKVNSETQFPAMRGCLTEKEV 180
Qy 181 A-HQMEIQLSYNFTLKQSSFVGWTHQRFVELVVDNIRYLSQSNATTVQHEVFNVN 239
Db 181 ARQOLEFEAEENALEPKSAGDWTWAFLELVVDNIRYLSQSNATTVQHEVFNVN 240
Qy 240 IVDSFYHPLVDVILTGIDITWASNPLPTSGDLONVLEDFSIWKNYNLNNRLQHDVAHLF 299

Db 241 IVDSMTYQQLGTYYILIGIBIWNQGNVFPMT-STEQVLNDFSQWKQISL-SQLQDAHMF 298
Qy 300 IKDTQGMKLGVAVKVIGICQNPENTGVDFEDNRLVVFATTLGHELGHNLGMQHDQWCV 359
Db 299 IKNSLISILGLAYVAGICRPPIDCGVDNFGQDWTSLFANTVAHELGTLMQHDDEFCFC 358
Qy 360 ELQWCIIMHAYRKVTTKFSNCSYAQYWDSTISSGLCTIOPPPYGNIFRLKYCGNLVVEEGE 419
Db 359 GERGCIMNTRVPAEFTNCSYADFMKTTTLNQSGSLNPPRLGEIFMLKRCGNGVVEREE 418
Qy 420 EDCGTTIRQCAKDPCCLLNCTLHPGAACAFGICCKCKFLPSGTLCRQQVGECDLPEWCN 479
Db 419 QDCGVSQCEQDACCLLNCTLRGAACAFGLCKCKFMPSGELCHQEVNECDLPEWCN 478
Qy 480 GTSHQCPDDVYVQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQSCYQEIINTQGN 539
Db 479 GTSHQCPEDRYVQDGIPCSDSAYCYQRCNNHDOHCREIFGDKAKSASQNCYKEINSQGN 538
Qy 540 RFHCGIGVTYVYKCTPDMCGRVQCEVGVIPNLIEHSTVQOQFHLNDTTCWGTDYHLG 599
Db 539 RFHCGINGITYLKCHISDVFCGRVQCEVNRDIPLLQDHFTLQHTHINGVTCWGDYHLR 598
Qy 600 MAIPDIEGVKDGTVCGPEKICIRKCKASVMHLSQACQPKTCNMRGICNNKHQCHCHNEWA 659
Db 599 MNISDIEGVKDGTVCGPGKICIKKCVLSVLSHVCLPETCNMKGICNNKHCHCHGCGWS 658
Qy 660 PPYCKDKYGGSDSGPPKNNMEGLNMGKRLYLSLLCLLPLVAFLLFCLHVLFFKRTK 719
Db 659 PPYQHRGYGGSIDSGPASAKRGVFLPLI-VIPSLSVLTFTVGLMYLRLQCSGPKETK 717
Qy 720 S 720
Db 718 A 718
RESULT 3
US-09-949-016-9212
; Sequence 9212, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9212
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9212
Query Match 52.3%; Score 2108.5; DB 4; Length 826;
Best Local Similarity 51.8%; Pred. No. 2.5e-178;
Matches 381; Conservative 112; Mismatches 217; Indels 25; Gaps 6;
Qy 5 EPLVHVRVTLILLWGMFLSISGHSQARPSQFTSPSEVVIPLKVISRGRGAKAPGWSYS 64
Db 2 EAFNKKMLLLHCLGFLVLSGSHIQDEHPQVHSPPDVVPIVRITGTTTGMTTPPGWLSVI 61
Qy 65 LRFGGQRYIVHMRVKNLLFAAHLPVFTYTEQHALLODQPFIODDCYHYGVGEVPESLVA 124
Db 62 LPFGQKHIIHKKVKKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYHYGVGEDEPSLVS 121

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 16:25:28 ; Search time 123 Seconds
(without alignments)
2282.830 Million cell updates/sec

Title: US-10-633-202-2

Perfect score: 4030

Sequence: 1 MAVGRLVIRVTLILLWFG.....LFLHLVLFKRTKSKEDSEG 726

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4030	100.0	726	2	AAY17413 Human SVP
2	4030	100.0	776	6	ABR39425 Human GEN
3	4004	99.4	726	2	AAY03223 Amino aci
4	4004	99.4	726	7	ADC78877 Human PRO
5	2435	60.4	722	2	AAY28655 Human SVP
6	2435	60.4	722	6	ABG76200 Human sna
7	2102.5	52.2	786	3	AAB07739 A snake v
8	2102.5	52.2	787	3	AAB07740 A snake v
9	2102.5	52.2	820	3	AAB07741 A snake v
10	2102.5	52.2	820	7	ADJ69446 Human hea
11	1717.5	42.6	542	2	AAY03224 Amino aci
12	1654	41.0	535	4	AAE13056 ADAM-20di
13	1646	40.8	535	2	AAY17414 SVPH1-26
14	1387.5	34.4	790	3	AAB07705 Human aci
15	1387.5	34.4	790	4	AAU12273 Human PRO
16	1387.5	34.4	790	6	ABO17717 Novel hum
17	1387.5	34.4	790	6	ABU80971 Human PRO
18	1387.5	34.4	790	6	ABU86671 Human PRO
19	1387.5	34.4	790	6	ABU59752 Novel sec
20	1387.5	34.4	790	6	ABO24942 Human sec
21	1387.5	34.4	790	6	ABU66947 Human sec
22	1387.5	34.4	790	6	ADA5723 Novel hum
23	1387.5	34.4	790	6	ADA76154 Human PRO
24	1387.5	34.4	790	6	ADA18804 Human PRO
25	1387.5	34.4	790	6	ADA61427 Homo sapi

26	1387.5	34.4	790	6	ADB19212 Novel hum
27	1387.5	34.4	790	6	ADB27753 Human PRO
28	1387.5	34.4	790	6	ADA86232 Novel hum
29	1387.5	34.4	790	6	ADB15796 Human PRO
30	1387.5	34.4	790	6	ADA47582 Human PRO
31	1387.5	34.4	790	6	ADA67377 Human PRO
32	1387.5	34.4	790	6	ADB30384 Human PRO
33	1387.5	34.4	790	6	ADA85680 Novel hum
34	1387.5	34.4	790	6	ADA96892 Human PRO
35	1387.5	34.4	790	6	ADA79196 Human PRO
36	1387.5	34.4	790	6	ADA87335 Novel hum
37	1387.5	34.4	790	6	ADB16537 Human PRO
38	1387.5	34.4	790	6	ADA91629 Novel hum
39	1387.5	34.4	790	6	ADB14692 Human PRO
40	1387.5	34.4	790	6	ADB18653 Novel hum
41	1387.5	34.4	790	6	ADA93868 Human PRO
42	1387.5	34.4	790	6	ADB19764 Novel hum
43	1387.5	34.4	790	6	ADB13076 Human PRO
44	1387.5	34.4	790	6	ABO43250 Novel hum
45	1387.5	34.4	790	6	ADA74330 Human PRO

ALIGNMENTS

RESULT 1

AAY17413

ID AAY17413 standard; protein; 726 AA.

XX AAY17413;

XX 26-JUL-1999 (first entry)

XX Human SVPH1-26 protein.

XX Human; SVPH1-26; proteinase; testis; fertilisation; spermatogenesis;

XX birth control; detergent additive; diagnosis; testicular cancer.

XX Homo sapiens.

XX WO9923228-A1.

XX 14-MAY-1999.

XX 30-OCT-1998; 98WO-US022965.

XX 30-OCT-1997; 97US-0063571P.

XX (IMMV) IMMUNEX CORP.

XX Cerretti DP;

XX WPI; 1999-337487/28.

XX N-PSDB; AAX56461.

XX New human SVPH1-26 DNA useful for the diagnosis and prognosis of testicular cancers.

XX Claim 2; Page 73-83; 96pp; English.

The present sequence is human SVPH1-26. Human SVPH1-26 polypeptides are proteinases implicated in fertilisation and spermatogenesis. They can be used as therapeutic agents. A proteinase inhibitor of the catalytic domain would inhibit SVPH1-26 activity and would be useful as a method for birth control. Also, an inhibitor of the disintegrin domain of SVPH1-26 may affect fertilisation. The proteinase activity of SVPH1-26 can also be used as a detergent additive for the removal of stains having a protein component. The SVPH1-26 polypeptides and fragments can also be used as molecular weight markers, as markers for determination of isoelectric points of sample proteins and as controls for establishing the extent of fragmentation of a protein sample. The products can also be used for identifying, separating or purifying cells that express SVPH1-26 polypeptides such as testis cells. They can be used for the diagnosis and

CC	prognosis of testicular cancers	
XX		
SQ	Sequence 726 AA;	
	Query Match 100.0%; Score 4030; DB 2; Length 726;	
	Best Local Similarity 100.0%; Pred. No. 1.9e-297;	
	Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MAVGEPLVHIRVTLTLLWFGMFLSISGHSQARPSQYFTSPVVIPLVKVISRGRGAKAPGW 60	
DB	1 MAVGEPLVHIRVTLTLLWFGMFLSISGHSQARPSQYFTSPVVIPLVKVISRGRGAKAPGW 60	
QY	61 LSYSLRFGGQRYIVHMRVNKLLFAAHLPVFTYTEQHALLODPFIQDDCYHYHGVGEVPE 120	
DB	61 LSYSLRFGGQRYIVHMRVNKLLFAAHLPVFTYTEQHALLODPFIQDDCYHYHGVGEVPE 120	
QY	121 SLVALSTCSGGFLGMLQINDLVYEIKPISVSATFEHLVVKIDSDDTQFPFMRGCLTEEKI 180	
DB	121 SLVALSTCSGGFLGMLQINDLVYEIKPISVSATFEHLVVKIDSDDTQFPFMRGCLTEEKI 180	
QY	181 AHQMEQLQSYNFTLKQSSFVGWTHQRFVVELVVVDNIRYLFQSQATTTVQHEVFNVNI 240	
DB	181 AHQMEQLQSYNFTLKQSSFVGWTHQRFVVELVVVDNIRYLFQSQATTTVQHEVFNVNI 240	
QY	241 VDSFYHPLEVDVILTGIDIWTSNPLTSGDLNVLDFSIWKYNLNLRQLQHDVAHLFI 300	
DB	241 VDSFYHPLEVDVILTGIDIWTSNPLTSGDLNVLDFSIWKYNLNLRQLQHDVAHLFI 300	
QY	301 KDTQGMKLGVAIVKIGIQNPFNTGVDVFEDNRLVVFATITLGHGHLNLMQHDQTCVCE 360	
DB	301 KDTQGMKLGVAIVKIGIQNPFNTGVDVFEDNRLVVFATITLGHGHLNLMQHDQTCVCE 360	
QY	361 LQWCMHAYRKVTTKFSNCSYAQYWDSTISSGLCIOPPPYPGNIFRLKYCGNLVVEEGE 420	
DB	361 LQWCMHAYRKVTTKFSNCSYAQYWDSTISSGLCIOPPPYPGNIFRLKYCGNLVVEEGE 420	
QY	421 CDCGTIRQCAKDPCCLLNCTLHPGAACAFGICCKCKFLPSGLTCRQQVGECDLPEWCNG 480	
DB	421 CDCGTIRQCAKDPCCLLNCTLHPGAACAFGICCKCKFLPSGLTCRQQVGECDLPEWCNG 480	
QY	481 TSHQCPDDVVQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQCYQINTQGNR 540	
DB	481 TSHQCPDDVVQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQCYQINTQGNR 540	
QY	541 FGHGIVGTTIVVKWTDPDMCGRVQCNVGVIPNLIEHSTVQQFHLNDTTCWGTDYHLGM 600	
DB	541 FGHGIVGTTIVVKWTDPDMCGRVQCNVGVIPNLIEHSTVQQFHLNDTTCWGTDYHLGM 600	
QY	601 AIPDIGVKDGTGCGPEKICIRKCKASMVHLSQACQPKTCNMKGICNNKQHCNHEWAP 660	
DB	601 AIPDIGVKDGTGCGPEKICIRKCKASMVHLSQACQPKTCNMKGICNNKQHCNHEWAP 660	
QY	661 PYCKDKYGGASGSGPPPKNMGLNVMGKRLYLSLLCLPLVAFLLFCLHVLFKKRTKS 720	
DB	661 PYCKDKYGGASGSGPPPKNMGLNVMGKRLYLSLLCLPLVAFLLFCLHVLFKKRTKS 720	
QY	721 KEDEEG 726	
DB	721 KEDEEG 726	
RESULT 2		
ABR39425		
ID	ABR39425 standard; protein; 776 AA.	
XX		
AC	ABR39425;	
XX		
DT	12-JUN-2003 (first entry)	
XX		
DE	Human GENSET polypeptide clone name v1ADAM20.	
XX		
KW	GENSET; cytostatic; gene therapy; cancer; transgenic; human.	
XX		

OS	Homo sapiens.	
XX		
PN	WO2003014151-A2.	
XX		
PD	20-FEB-2003.	
XX		
PF	15-OCT-2001; 2001WO-IB002321.	
PR	10-AUG-2001; 2001US-03111305P.	
PR	24-AUG-2001; 2001US-0314734P.	
PR	07-SEP-2001; 2001US-0318204P.	
PR	01-OCT-2001; 2001US-0326470P.	
XX		
PA	(GEST) GENSET SA.	
XX		
PI	Bejanin S, Tanaka H;	
XX		
DR	WPI; 2003-256539/25.	
DR	N-PSDB; AB276248.	
XX		
PT	New GENSET gene, useful for preparing a composition for treating GENSET-related disorders.	
XX		
PS	Claim 2; Page 250-252; 301pp; English.	
XX		
CC	The invention relates to isolated GENSET polynucleotides and encoded polypeptides. The GENSET gene is useful for preparing a composition for treating GENSET-related disorders e.g., cancer. Sequences ABR39423-448 represent the novel GENSET polypeptide sequences	
XX		
SQ	Sequence 776 AA;	
	Query Match 100.0%; Score 4030; DB 6; Length 776;	
	Best Local Similarity 100.0%; Pred. No. 2.1e-297;	
	Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MAVGEPLVHIRVTLTLLWFGMFLSISGHSQARPSQYFTSPVVIPLVKVISRGRGAKAPGW 60	
DB	51 MAVGEPLVHIRVTLTLLWFGMFLSISGHSQARPSQYFTSPVVIPLVKVISRGRGAKAPGW 110	
QY	61 LSYSLRFGGQRYIVHMRVNKLLFAAHLPVFTYTEQHALLODPFIQDDCYHYHGVGEVPE 120	
DB	111 LSYSLRFGGQRYIVHMRVNKLLFAAHLPVFTYTEQHALLODPFIQDDCYHYHGVGEVPE 170	
QY	121 SLVALSTCSGGFLGMLQINDLVYEIKPISVSATFEHLVVKIDSDDTQFPFMRGCLTEEKI 180	
DB	171 SLVALSTCSGGFLGMLQINDLVYEIKPISVSATFEHLVVKIDSDDTQFPFMRGCLTEEKI 230	
QY	181 AHQMEQLQSYNFTLKQSSFVGWTHQRFVVELVVVDNIRYLFQSQATTTVQHEVFNVNI 240	
DB	231 AHQMEQLQSYNFTLKQSSFVGWTHQRFVVELVVVDNIRYLFQSQATTTVQHEVFNVNI 290	
QY	241 VDSFYHPLEVDVILTGIDIWTSNPLTSGDLNVLDFSIWKYNLNLRQLQHDVAHLFI 300	
DB	291 VDSFYHPLEVDVILTGIDIWTSNPLTSGDLNVLDFSIWKYNLNLRQLQHDVAHLFI 350	
QY	301 KDTQGMKLGVAIVKIGIQNPFNTGVDVFEDNRLVVFATITLGHGHLNLMQHDQTCVCE 360	
DB	351 KDTQGMKLGVAIVKIGIQNPFNTGVDVFEDNRLVVFATITLGHGHLNLMQHDQTCVCE 410	
QY	361 LQWCMHAYRKVTTKFSNCSYAQYWDSTISSGLCIOPPPYPGNIFRLKYCGNLVVEEGE 420	
DB	411 LQWCMHAYRKVTTKFSNCSYAQYWDSTISSGLCIOPPPYPGNIFRLKYCGNLVVEEGE 470	
QY	421 CDCGTIRQCAKDPCCLLNCTLHPGAACAFGICCKCKFLPSGLTCRQQVGECDLPEWCNG 480	
DB	471 CDCGTIRQCAKDPCCLLNCTLHPGAACAFGICCKCKFLPSGLTCRQQVGECDLPEWCNG 530	
QY	481 TSHQCPDDVVQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQCYQINTQGNR 540	
DB	531 TSHQCPDDVVQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQCYQINTQGNR 590	
QY	541 FGHGIVGTTIVVKWTDPDMCGRVQCNVGVIPNLIEHSTVQQFHLNDTTCWGTDYHLGM 600	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 16:31:09 ; Search time 124 Seconds
(without alignments)
2998.141 Million cell updates/sec

Title: US-10-633-202-2

Perfect score: 4030

Sequence: 1 NAVGELPLVHVRVTLMLWFG.....LFCLHVLFKKRTKSKDEEG 726

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	DB	ID	Description
1	4004	99.4	726	1	AD20	HUMAN	Q43506 homo sapien
2	2435	60.4	722	1	AD21	HUMAN	Q9ukj8 homo sapien
3	2102.5	52.2	820	1	AD29	HUMAN	Q9ukf5 homo sapien
4	2050	50.9	729	1	AD21	MOUSE	Q9ij76 mus musculus
5	1994.5	49.5	760	1	AD35	MOUSE	Q9r159 mus musculus
6	1953.5	48.5	756	2	Q7M762		Q7m762 mus musculus
7	1946	48.3	760	2	Q7M763		Q7m763 mus musculus
8	1934	48.0	763	2	Q811Q4		Q811q4 mus musculus
9	1899.5	47.1	761	2	Q8CDV3		Q8cdv3 mus musculus
10	1892.5	47.0	761	1	AD24	MOUSE	Q9r160 mus musculus
11	1780.5	44.2	714	2	Q8K4K0		Q8k4k0 mus musculus
12	1774	44.0	713	2	Q7M765		Q7m765 mus musculus
13	1772.5	44.0	702	2	Q8BMR5		Q8bmr5 mus musculus
14	1759.5	43.7	736	2	Q7M766		Q7m766 mus musculus
15	1758	43.6	699	2	Q6IMH6		Q6imh6 mus musculus
16	1707.5	42.4	697	2	Q8BMR4		Q8bmr4 mus musculus
17	1697.5	42.1	697	1	AD26	MOUSE	Q9r158 mus musculus
18	1387.5	34.4	790	1	AD30	HUMAN	Q9ukf2 homo sapien
19	1387.5	34.4	790	2	Q8TBZ7		Q8tbz7 homo sapien
20	1338.5	33.2	819	1	AD09	HUMAN	Q13443 homo sapien
21	1333.5	33.1	775	2	Q76Kt5		Q76kt5 gallus gall
22	1332	33.1	873	2	Q42595		Q42595 xenopus lae
23	1328.5	33.0	845	1	AD09	MOUSE	Q61072 mus musculus
24	1300.5	32.3	732	2	Q811Q3		Q811q3 mus musculus
25	1250.5	31.0	838	2	O19056		O19056 papio anubi
26	1237.5	30.7	905	2	Q28476		Q28476 macaca fasc
27	1231	30.5	825	2	O46652		O46652 papio anubi
28	1230.5	30.5	821	2	O19060		O19060 saguinus oe
29	1229.5	30.5	836	2	O19057		O19057 pongo pygma
30	1226	30.4	825	2	Q28477		Q28477 macaca fasc
31	1215	30.1	792	2	O19061		O19061 saguinus oe

32	1210.5	30.0	789	2	Q66HK9		Q66hk9 rattus norv
33	1208.5	30.0	789	1	AD01	RAT	P70505 rattus norv
34	1204	29.9	791	1	AD1A	MOUSE	Q60813 mus musculus
35	1193	29.6	812	2	Q77779		O77779 bos taurus
36	1193	29.6	919	2	Q28659		Q28659 oryctolagus
37	1190.5	29.5	763	2	Q8CGQ2		Q8cgq2 mus musculus
38	1170.5	29.0	713	2	Q28485		Q28485 macaca fasc
39	1169.5	29.0	703	2	Q8C0V3		Q8cdv3 mus musculus
40	1168.5	29.0	756	2	Q6IMH7		Q6imh7 mus musculus
41	1163.5	28.9	730	2	O19051		O19051 oryctolagus
42	1160.5	28.8	754	2	Q8CGQ1		Q8cgq1 mus musculus
43	1153.5	28.6	732	2	Q28484		Q28484 macaca fasc
44	1152.5	28.6	751	2	P70535		P70535 rattus norv
45	1149	28.5	751	2	Q6IMH8		Q6imh8 mus musculus

ALIGNMENTS

RESULT 1
AD20 HUMAN
ID AD20 HUMAN STANDARD; PRT; 726 AA.
AC Q43506: Q9UKJ9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 25-OCT-2004 (Rel. 45, Last annotation update)
DE ADAM 20 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 20).
GN Name=ADAM20;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=98137801; PubMed=9469942; DOI=10.1016/S0378-1119(97)00597-0;
RA "Hoof van Huijsduijn R.;
RT "ADAM 20 and 21; two novel human testis-specific membrane metalloproteases with similarity to fertilin-alpha.";
RL Gene 206:273-282(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99453762; PubMed=10524237; DOI=10.1016/S0378-1119(99)00302-9;
RA Poindexter K., Nelson N., DuBose R.F., Black R.A., Cerretti D.P.;
RT "The identification of seven metalloproteinase-disintegrin (ADAM) genes from genomic libraries.";
RL Gene 237:61-70(1999).
CC -!- FUNCTION: May be involved in sperm maturation and/or fertilization.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (Potential).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Testis specific.
CC -!- DOMAIN: A tripeptide motif (VGE) within disintegrin-like domain could be involved in the binding to egg integrin receptor and thus could mediate sperm/egg binding.
CC -!- DOMAIN: The cysteine-rich domain encodes putative cell-fusion peptides, which could be involved in sperm-egg fusion.
CC -!- PTM: Has no obvious cleavage site for furin endopeptidase, suggesting that the proteolytic processing is regulated.
CC -!- MISCELLANEOUS: May be the functional equivalent of ADAM 1/fertilin alpha which is a pseudogene in human.
CC -!- SIMILARITY: Belongs to the peptidase M12B family.
CC -!- SIMILARITY: Contains 1 disintegrin domain.
CC -!- SIMILARITY: Contains 1 EGF-like domain.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

```

CC -----
DR ENBL; AF029899; AAC52041.1; -.
DR EMBL; AF158643; AAD52524.1; -.
DR HSSP; P18619; IFVL.
DR MEROPS; M12.218; -.
DR GO; GO:0008237; F-metalloproteinase activity; TAS.
DR GO; GO:0007338; P-fertilization (sensu Animalia); TAS.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR001818; Pept M10A_M12B.
DR InterPro; IPR006025; Pept M Zn BS.
DR InterPro; IPR001590; Peptidase_M12B.
DR InterPro; IPR002870; Peptidase_M12B_N.
DR Pfam; PF00200; Disintegrin; 1.
DR Pfam; PF01562; Pep M12B propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR ProDom; PD000664; Disintegrin; 1.
DR PROSITE; PS02015; ADAM_MBP; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; FALSE NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
DR PROSITE; PS00214; DISINTEGRIN_2; 1.
DR PROSITE; PS00022; EGF_1; FALSE NEG.
DR PROSITE; PS01186; EGF_2; FALSE NEG.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW EGF-like domain; Glycoprotein; Hydrolase; Metalloprotease; Signal;
KW TRANSMEMBRANE; Zinc; Zymogen.
FT SIGNAL 1 31 Potential.
FT PROPEP 32 206 Potential.
FT CHAIN 207 726 ADAM 20.
FT DOMAIN 207 693 Extracellular (Potential).
FT TRANSMEM 694 714 Potential.
FT DOMAIN 715 726 Cytoplasmic (Potential).
FT DOMAIN 207 384 Metalloprotease.
FT DOMAIN 407 493 Disintegrin-like.
FT DOMAIN 494 634 Cys-rich.
FT DOMAIN 635 663 EGF-like.
FT SITE 173 173 Cysteine switch.
FT ACT_SITE 343 343 Potential.
FT METAL 342 342 Zinc (catalytic) (Potential).
FT METAL 346 346 Zinc (catalytic) (Potential).
FT METAL 352 352 Zinc (catalytic) (Potential).
FT DISULFID 317 394 By similarity.
FT DISULFID 357 379 By similarity.
FT DISULFID 465 478 Potential.
FT DISULFID 635 646 By similarity.
FT DISULFID 640 652 By similarity.
FT DISULFID 654 663 By similarity.
FT CARBOHYD 191 191 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 226 226 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 378 378 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 438 438 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 479 479 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 587 587 N-linked (GlcNAc... ) (Potential).
FT CONFLICT 19 19 L > P (in Ref. 2).
FT CONFLICT 109 109 W > C (in Ref. 2).
FT CONFLICT 637 637 R -> P (in Ref. 2).
SQ SEQUENCE 726 AA; 81710 MW; BF356A6771CD1C12 CRC64;

Query Match 99.4%; Score 4004; DB 1; Length 726;
Best Local Similarity 99.6%; Pred. No. 1.7e-281;
Matches 723; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVGEPLVHVRVTLVLLWFGNFIISGHSQARSQYFTSPSEVVIPLKVISRGKAPGW 60
DB |||||||
DB 1 MAVGEPLVHVRVTLVLLWFGNFIISGHSQARSQYFTSPSEVVIPLKVISRGKAPGW 60
QY 61 LSYSLRFGGORYIVHMRVKNLLFAAHLPVFTYTEQHALLODQPFIODDCYHYGVGVPE 120
DB |||||||
DB 61 LSYSLRFGGORYIVHMRVKNLLFAAHLPVFTYTEQHALLODQPFIODDCYHYGVGVPE 120

```

```

QY 121 SLVALSTCGSGFLGMLQINDLVYEIKPIVSATFEHLVYKIDSDDTQFPFMRGLTEEKI 180
DB |||||||
DB 121 SLVALSTCGSGFLGMLQINDLVYEIKPIVSATFEHLVYKIDSDDTQFPFMRGLTEEKI 180
QY 181 AHOMELQLSYNFTLTKOSSFVGWTHQRFVELVVDNIRYLFSQSNAITVQHEVFVNVI 240
DB |||||||
DB 181 AHOMELQLSYNFTLTKOSSFVGWTHQRFVELVVDNIRYLFSQSNAITVQHEVFVNVI 240
QY 241 VDSFYHPLEVDVILTGIDITWASNPLTSGDDLNVLEDFSIWKNNLNRLQHDVAHLFI 300
DB |||||||
DB 241 VDSFYHPLEVDVILTGIDITWASNPLTSGDDLNVLEDFSIWKNNLNRLQHDVAHLFI 300
QY 301 KDTQGMKLGVAVYKGIQNPFTGVDFEDNRLVWFAITLGHGHLNGLQWHDTCVCE 360
DB |||||||
DB 301 KDTQGMKLGVAVYKGIQNPFTGVDFEDNRLVWFAITLGHGHLNGLQWHDTCVCE 360
QY 361 LQWCIMHAYRKVTTKFSNCSYAQYWDSTISSGLCIQPPYPGNIPLKCYCGNLVVEGEE 420
DB |||||||
DB 361 LQWCIMHAYRKVTTKFSNCSYAQYWDSTISSGLCIQPPYPGNIPLKCYCGNLVVEGEE 420
QY 421 CDGCTIROCAKDPCCLLNCTLHPGACAFGICCKCKFLPSGTLCHQOYGECDLPWCNG 480
DB |||||||
DB 421 CDGCTIROCAKDPCCLLNCTLHPGACAFGICCKCKFLPSGTLCHQOYGECDLPWCNG 480
QY 481 TSHQCPDDVYVQDGISSCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQCYQINTQGNR 540
DB |||||||
DB 481 TSHQCPDDVYVQDGISSCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQCYQINTQGNR 540
QY 541 FGHGIVGTYYVKCWTPTDIMCRVQCVENGVIPNLIEHSTVQOQFHLNDDTCWGTDFHLM 600
DB |||||||
DB 541 FGHGIVGTYYVKCWTPTDIMCRVQCVENGVIPNLIEHSTVQOQFHLNDDTCWGTDFHLM 600
QY 601 ATPDGEVVDGTVCGPEKICIRKCKASVHLSQACOPKTCNMRGICNNKQHCNHEWAP 660
DB |||||||
DB 601 ATPDGEVVDGTVCGPEKICIRKCKASVHLSQACOPKTCNMRGICNNKQHCNHEWAP 660
QY 661 PYCKDKGYGGSADSGPPPKNNMEGLNVMGKLYLSLLCLPLVAFLLFCLHLVFKKRTKS 720
DB |||||||
DB 661 PYCKDKGYGGSADSGPPPKNNMEGLNVMGKLYLSLLCLPLVAFLLFCLHLVFKKRTKS 720
QY 721 KEDEEG 726
DB |||||||
DB 721 KEDEEG 726

RESULT 2
AD21_HUMAN STANDARD; PRT; 722 AA.
ID AD21_HUMAN
AC Q9UKJ8; O43507;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE ADAM 21 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 21).
GN Name=ADAM21;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99453762; PubMed=10524237; DOI=10.1016/S0378-1119(99)00302-9;
RA Poindexter K., Nelson N., DuBoise R.F., Black R.A., Cerretti D.P.;
RT "The identification of seven metalloproteinase-disintegrin (ADAM) genes from genomic libraries.";
RL Gene 237:61-70 (1999).
RN [2]
RP SEQUENCE OF 181-722 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=98137801; PubMed=9469942; DOI=10.1016/S0378-1119(97)00597-0;
RA Hoof van Huijdeduyn R.;
RT "ADAM 20 and 21; two novel human testis-specific membrane metalloproteases with similarity to fertilin-alpha.";
RT

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 18:29:15 ; Search time 51 Seconds
(without alignments)
1369.674 Million cell updates/sec

Title: US-10-633-202-2
Perfect score: 4030
Sequence: 1 MAVGRLVHIVTLLLLMFG.....LFCLHVLFKRTKSEDEEG 726
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues 283416
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1237.5	30.7	825	2 S55059	fertilin alpha-I -
2	1226	30.4	905	2 S55060	fertilin alpha-II
3	1186	29.4	655	2 JC7850	disintegrin and me
4	1170.5	29.0	713	2 I65253	disintegrin-like t
5	1153.5	28.6	732	2 I52361	testicular metallo
6	1136	28.2	660	2 S71949	metalloproteinase
7	1109.5	27.5	735	2 I48101	ADAM 6 protein pre
8	991.5	24.6	600	2 I49281	fertilin alpha pre
9	987.5	24.5	903	2 S60257	meltrin alpha - mo
10	918.5	22.8	756	2 S47656	tMDC II protein -
11	910	22.6	734	2 JC4861	fertilin beta cha
12	896.5	22.2	735	2 G02937	fertilin beta - cr
13	868	21.5	736	2 S47645	tMDC I protein - c
14	856	21.2	777	2 I48100	ADAM 5 protein pre
15	852.5	21.2	609	2 S55270	catrocollastatin p
16	829	20.6	473	2 I49283	ADAM 4 protein pre
17	826	20.5	610	2 JC7530	vascular apoptosis
18	818.5	20.3	617	2 S48160	metalloproteinase
19	816	20.2	616	2 A55796	ecarin precursor -
20	815.5	20.2	571	2 S24789	jararagin C precu
21	815	20.2	823	2 S18968	cytostatin precurs
22	788	19.6	610	2 JC8056	halsase - Gloydiu
23	786	19.5	814	2 G02390	disintegrin-like m
24	784	19.5	776	2 S28258	androgen-regulated
25	782.5	19.4	789	2 S28259	androgen-regulated
26	768.5	19.1	826	2 A60385	monocyte surface a
27	739.5	18.3	549	2 S48169	metalloproteinase
28	724.5	18.0	670	2 I65967	disintegrin-like m
29	712.5	17.7	952	2 T18900	disintegrin and me

30 677.5 15.8 419 2 A59414 metalloproteinase
31 641.5 15.9 429 2 A42972 coagulation factor
32 641 15.9 478 2 A43296 atrolysin E (EC 3.
33 617 15.3 484 2 JC8020 metalloproteinase-
34 605.5 15.0 419 2 S41607 atrolysin A (EC 3.
35 602 14.9 1042 2 T26644 hypothetical prote
36 592 14.7 480 1 A30065 trigranin precurs
37 580.5 14.4 524 2 S38539 disintegrin-like m
38 577 14.3 416 2 A37877 hemorrhagic protel
39 576.5 14.3 481 2 S43125 trimucin precursor
40 568 14.1 357 2 S23403 sperm surface prot
41 555 13.8 481 2 JC4342 fibrinolytic prote
42 553 13.7 478 2 JQ1301 hemorrhagic protei
43 551 13.7 478 2 JC4880 fibrinolytic metal
44 492 12.2 407 2 S6260 metalloproteinase
45 476 11.8 411 1 HYSNFA fibrolase (EC 3.4.

ALIGNMENTS

RESULT 1

S55059
fertilin alpha-I - crab-eating macaque
C:Species: Macaca fascicularis (crab-eating macaque)
C>Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
C:Accession: S55059
R:Perry, A.C.F.; Gichuhi, P.M.; Jones, R.; Hall, L.
Biochem. J. 307, 843-850, 1995
A:Title: Cloning and analysis of monkey fertilin reveals novel alpha subunit isoforms.
A:Reference number: S55059; MUID:95260313; PMID:7741716
A:Accession: S55059
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-905 <PER>
A:Cross-references: UNIPROT:Q28476; EMBL:X79808; NID:G794072; PIDN:CAA56203.1; PID:G79407;
C:Superfamily: mouse meltrin alpha; disintegrin homology
F:442-522/Domain: disintegrin homology <Dis>
F:376/Active site: Glu #status predicted

Query Match 30.7%; Score 1237.5; DB 2; Length 905;
Best Local Similarity 36.7%; Pred. No. 8.5e-78;
Matches 269; Conservative 121; Mismatches 284; Indels 59; Gaps 19;

QY	10	IRVTLLLMFGMFLSIGHSQARPSQYFTSPVVIPLKVISRGAKAPGMLSYSLRFGG	69
DB	52	IRLEILLALLVIFVPSMHCH---LGSIIYYSFYBIIIPKRLTVQGGDSSVEG-LSYLLFMQG	107
QY	70	QRYIVHMRVKNLLFAAHLPVFTYITQHALLOQDPFIQDDCYHYGVGVPSLSVALSTCS	129
DB	108	QKHLVHLVKRSHFVNNFPVYSY-HNGILGQSPFISHDCHYEGYIEGVSGSVSVNTCA	166
QY	130	GGFLQMLQINDLAVYIKPISVSATFEHLVYKIDSDDTQPPMRCGLTE-----	177
DB	167	-GLRGILLKKEKSYSEIPMDSRRRFEHLVYTNABH-----ARVSCGVTSRDSHVSTSWQ	221
QY	178	-EKIAHQMELQSLSYNFTLKQSSFVGWTHQRFVELVWVVDNIRYLFQSNATTQVEFVN	236
DB	222	GSRKPHDLQ-ALSY-----LWSHTKYVENFVVVNNQRFQWMSGNINETVQRVVD	269
QY	237	VNVIVDSFYHPLLEVIVLTGDIWTASNPPLPSGDLNDVLEDFSIWKNYNLNRRLOHVA	296
DB	270	VIALANGFTRGINTENVLAGMEIWTEGLDIDVAVDLQITLRNFRWRQMLFERRAKHVA	329
QY	297	HLFIKDTQGMKLGVAVKGIQNPENTGVDPEDNRLVVFAITLGHLEHNLGMQHDQW	356
DB	330	HMIVGHPQNTGQAFLSGACSSGFAAVESPHHEMLLFAALMVHLEHNLGIQHDHSA	389
QY	357	CVC-BLOWCIMHAYRKVTTKFSNCSSYAQWYDSTI--SSGLCTIQPPYP-GNIFRLKYCN	412
DB	390	CFCREKHFLCMENITKSGFSNCS-SDYFHQFLREHKGACLFNKRPRGRKRDSACGN	448
QY	413	LVBSEGECDCTIRQCAKDPCCLLNCTLHPGAACAAGICCKDKFLPSGTLCRQOVGEC	472

```
Db 449 GWVEDTEEDCGS--ACHLDPCDDPTCTLKEHAECSHGLCCLDCTFRRKGFCLCRPTQDEC 506
Qy 473 DLPEWNGTSHOCDDVYVQDGISCNVAFCEYKTCNNHDIQCKEIPGQDARSASQCYQ 532
Db 507 DLPEYCDGSSAECPADSYKQDGLTCDRIHYCSGGCKNPDQCNQCNVYGYPARSAPEDCYI 566
Qy 533 EINTQGNRFHCGIV---GTTYVKWTPDIMGRCVOCENGVIPNL-IEHSTVQOQFHLND 588
Db 567 SMNTRGRDFGNCGHPTEDQOQTYVTCSDDNVFCGKLICTGVQSLPRVKAQHTVTVQVPHDND 626
Qy 589 TTCGTGTYHLGMAIPDGEVKDGTVCGEPIKIRKCKASVMHLSQACQP-KTCNMGRGICN 647
Db 627 -WCMSMDADNITDPDNGNVHVGTSAPNKCCTDYSVHHSILLYDCRPEESCHGKVCN 685
Qy 648 NKQCHCHNHWAPPYCKDKGYGGSADSGPP---PKNMW-----GLNVMGKLYLSL 696
Db 686 NLRHCHCESGAPPDCKNPGNGSVDSGPPQMVTNNSSESGSISARGQSLRDQVDYKLV 745
Qy 697 LCLPLVAFLLFC 709
Db 746 VLLVPLFLVLLC 758

RESULT 2
S55060
fertilin alpha-II - crab-eating macaque
C:Species: Macaca fascicularis (crab-eating macaque)
C>Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
C:Accession: S55060
R:Perry, A.C.F.; Gichuhi, P.M.; Jones, R.; Hall, L.
Biochem. J. 307, 843-850, 1995
A:Title: Cloning and analysis of monkey fertilin reveals novel alpha subunit isoforms.
A:Reference number: S55059; MUID:95260313; PMID:7741716
A:Accession: S55060
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-825 <PER>
A:Cross-references: UNIPROT:Q28477; EMBL:X79809; NID:g794074; PIDN:CAA56204.1; PID:g794074
C:Superfamily: mouse meltrin alpha; disintegrin homology
F:443-523/Domain: disintegrin homology <DIS>
F:377/Active site: Glu #status predicted

Query Match 30.4%; Score 1226; DB 2; Length 825;
Best Local Similarity 35.8%; Pred. No. 4.8e-77;
Matches 275; Conservative 126; Mismatches 288; Indels 80; Gaps 20;

Qy 10 IRVTLILLWFMFLSISGHSGQARPSQYFTPEVVIPLKVISRGRG---AKAPGWLVSYSLR 66
Db 52 VRLGMLLLGILFLPSICCDMA--SVYSSVEIIPKRLMVRGSEDSVEKA---TYLLL 105
Qy 67 FGGORYIVHMRVKNLLFAAHLPVFTYTEQHALLQDQPFIODDCYHYGVGVPSLVALS 126
Db 106 MQQGHLLVHLKVKSHVFNFPVYSY-HNGILGQSPFISHDCHYEYIEGVSGSFVSVN 164
Qy 127 TCSGGFLGMLQINDLVYEIKPISVSATFEHLVYKIDSDDTQFPFMRGCLTE----- 177
Db 165 TCA-GLRGILIKEKKSYSIEPMSRRPEHVLVYMAHE-----ARVSGVTSRSHVVSTS 219
Qy 178 ----EKIAHQMELQSLSYFTLKQSSFFVGWTHQRFVELVVDNIRYLFQSNATTVQHE 233
Db 220 WQGSRRPHDLQ-ALSY-----LWSHTKYVEMFVVNNQRFQWGSNINETVQR 267
Qy 234 VFNVVNVDSPYHPLEVDVILTGIDITWASNPFTSGDLNVLDFSIWKYNLNLRLOH 293
Db 268 VVDVIALANGTRGINTVEVLAGEIWIETGDLIDVAVDLQITLRNFNRWQEMLFRAKH 327
Qy 294 DVAHLFTKDTQGMKGLVAYVYKGIQNPFTGVDVDFEDNRLVVFVAILTGLHGLHNLGMOHD 353
Db 328 DVAMIVGHHPQNTQGAFLSGACSSGFAAVSFHHEDMLLPALMVMHGLHNLGLOHD 387
Qy 354 TQWCVC-ELQWCIMHAYRKVTKFSNCSYAQYWDSTI--SSGLCIQPPYP-GNIFPKY 409
```

```
Db 388 HSACFCKREKHFCLMHENITKESGFSNCS-SDYFHQFLREHKGACLFNKPFRPRGRKRDSA 446
Qy 410 CGNLVVEEBECDCGTIROCAKDPCCLLNCTLHPGACAFGICKCKDCKPLPSGTLCRQOV 469
Db 447 CGNGVEDTEEDCGS--ACHLDPCDDPTCTLKEHAECSHGLCCLDCTFRRKGFCLCRPTQ 504
Qy 470 GECDLPEWNGTSHOCDDVYVQDGISCNVAFCEYKTCNNHDIQCKEIPGQDARSASQS 529
Db 505 DECDLPEYCDGSSAECPADSYKQDGLTCDRIHYCSGGCKNPDQCNQCNVYGYPARSAPED 564
Qy 530 CYOEINTQGNRFHCGIV---GTTYVKWTPDIMGRCVOCENGVIPNL-IEHSTVQOQFHL 585
Db 565 CYTSMNTRGRDFGNCGHPTEDQOQTYVTCSDDNVFCGKLICTGVQSLPRVKAQHTVTVQVPH 624
Qy 586 LNDTTGWTGTYHLGMAIPDGEVKDGTVCGEPIKIRKCKASVMHLSQACQP-KTCNMGRG 644
Db 625 DND-WCMSMDADNITDPDNGNVHVGTSAPNKCCTDYSVHHSILLYDCRPEESCHGK 683
Qy 645 ICNNKQCHCHNHWAPPYCKDKGYGGSADSGPPPKNMMEGLN----- 686
Db 684 VCNLRHCHCGSGFAPPDCKNPGNGSVDSGPAKPSDEIISREENRNHIVGHGNPQRGD 743
Qy 687 -----VMGKLYLSLCLLPLVAFLLFCLHVLFKR--TKSKEDEEG 726
Db 744 VSNKNKSLGKLVYIVPLLLALLFAGLIILASMGARKEILQRSQGDTEG 792

RESULT 3
JC7850
disintegrin and metalloproteinase (ADAM) 9 protein, short form - human
N:Alternate names: MDC9 protein; meltrin gamma
C:Species: Homo sapiens (man)
C>Date: 18-Nov-2002 #sequence_revision 18-Nov-2002 #text_change 09-Jul-2004
C:Accession: JC7850
R:Hotoda, N.; Koike, H.; Sasagawa, N.; Ishiura, S.
Biochem. Biophys. Res. Commun. 293, 800-805, 2002
A:Title: A secreted form of human ADAM9 has an alpha-secretase activity for APP.
A:Reference number: JC7850; MUID:22050095; PMID:12054541
A:Accession: JC7850
A:Molecule type: mRNA
A:Residues: 1-655 <HOT>
A:Cross-references: UNIPROT:Q8NFM6; GB:AF495383
C:Comment: This protein, which is a member of the a disintegrin and metalloprotease (ADAM)
lular functions. It is proteolytically active, and has an alpha-secretase activity for an
C:Genetics:
A:Gene: adam9s

Query Match 29.4%; Score 1186; DB 2; Length 655;
Best Local Similarity 37.5%; Pred. No. 2.2e-74;
Matches 249; Conservative 116; Mismatches 249; Indels 50; Gaps 15;

Qy 10 IRVTLILLWFMFLSISGHSGQARP-----SQYFTSPVVIPLKVISRGRGAKAP--GWLVS 63
Db 12 LRVYRLL-LGLVGPVLG--AARPGFQQTSHLSSYEITPWRLTRERREAPREYSKQVSY 68
Qy 64 SLRFSGORYIVHMRVKNLLFAAHLPVFTYTEQHALLQDQPFIODDCYHYGVGVPSLVL 123
Db 69 VIOAQEKHIIHLERNKDLPLDPFVYVYTNKGETLITDHPNIQNHCHYGYVEGVNSSI 128
Qy 124 ALSTCSGGFLGMLQINDLVYEIKPISVSATFEHLVYKIDSDDTQFPFMRGCLT----- 176
Db 129 ALSDCF-GLRGLLHLENASYGIEPLQNSSHFEHIIYRM--DDVYKEPLKCGVSNKDIEKE 185
Qy 177 -----BEKTAHQMELQSLSYFTLKQSSFFVGWTHQRFVELVVDNIRYLFQSNATTVQ 231
Db 186 TAKDEEBEPPSTQLRRRAVLPQT-----RYVELFIVVDKERYDMGRNQTAVR 236
Qy 232 HEFNVVNVDSYPYHPLEVDVILTGIDITWASNPFTSGDLNVLDFSIWKYNLNLRNL 291
Db 237 EEMILANVLDSDNYIMNIRIVLVGLEIWTNGNLINIVGAGDVLGNFVQWREKFLITRR 296
Qy 292 QHDVAHLFTKDTQGMKGLVAYVYKGIQNPFTGVDVDFEDNRLVVFVAILTGLHGLHNLGMO 351
```